Visualization of object detection data

Anna Lissborg - Oskar Jönsson

DEPARTMENT OF DESIGN SCIENCES
FACULTY OF ENGINEERING LTH | LUND UNIVERSITY
2017

MASTER THESIS

Visualization of object detection data

Anna Lissborg - Oskar Jönsson

Master thesis 2017

LUND
UNIVERSITY

Department of Design Sciences

MSc Thesis
ISRN LUTFD2/TFRT--1337--SE
ISSN 0280–5316

Department of Design Sciences
Faculty of Engineering LTH, Lund University
Box 118
SE-221 00 LUND
Sweden

Division: Ergonomics and Aerosol Technology
Subject: Interaction Design (MAMM01)
Supervisor: Johanna Persson, JOHANNA.PERSSON@DESIGN.LTH.SE
Supervisor: Mazdak Farzone, MAZDAK@AXIS.COM
Examiner: Joakim Eriksson, JOAKIM.ERIKSSON@DESIGN.LTH.SE

Printed in Sweden.
Lund 2017

Abstract

As a part of the innovation and growth within the Axis, object detection with new technology is explored. The car industry is lowering the costs on RAdio Detection And Ranging (RADAR) chipsets, making them a viable hardware to incorporate within a surveillance camera device. How can the new hardware improve surveillance and provide the end user with an enhanced experience? One challenge is to graphically visualize video content; triggered by object detection, with a user experience focus, which is the aim of the thesis.

Furthermore, the thesis performs a user-centered design approach without direct access to the end user, this enables a faster design process but the outcome may lack in features and design choices requested by an end user. An iterative design method was applied and separated the process into five phases. Phase one to three focused on developing and designing. In the forth phase, the outcome is validated and lastly, in phase five, the result is presented, discussed and concluded.

Keywords: User-centered design, UCD, none interactive, graphical visualization, video surveillance, object detection.

Sammanfattning

Som en del av innovationsarbetet och tillväxten inom Axis, har rörelsedetektering med hjälp av ny teknik utforskats. Genom sin utveckling har, bilindustrin sänkt kostnaderna på RAdio Detection And Ranging (RADAR) sensorer, vilket gör dem till en attraktiv hårdvara att inkludera i övervakningskameror och dess system. Men, hur kan den nya hårdvaran förbättra övervakning och ge slutanvändaren en bättre upplevelse? En av utmaningarna är att grafiskt visualisera videoinnehållet; triggat av rörelsedetektering, med användarupplevelsen i fokus.

Den användarcentrerad designprocess utan direkt tillgång till slutanvändaren, som utgör basen för detta arbete, möjliggör en snabbare designprocess. Resultatet kan dock sakna funktioner eller brista i design om slutanvändaren inte involveras i tidigt skede. En iterativ designprocess tillämpades och delade upp processen i fem steg. Steg ett till tre fokuserade på att ta fram funktionalitet, utveckla och designa. I det fjärde steget, validerades resultatet och slutligen i steg fem, presenterades och diskuterades resultatet.

Nyckelord: Användarcentrerad design, UCD, utan interaktion, grafisk visualisering, videoövervakning, rörelsedetektering.

Acknowledgements

Firstly, we would like to express our gratitude to Axis Communications AB and especially the PCNI team, for making us feel as a part of the team, and for giving us the opportunity to execute this master's thesis.

To our supervisors Mazdak Farzone at Axis Communications and Johanna Persson at the department of Design Sciences, we send our special thanks, they gave us the opportunity develop and supported us to in accomplishing our goal. Lastly, we would like to thank our co-workers and friends who helped us a lot during the process and always kept a good spirit in the office.

Lund, March, 2017.
Anna Lissborg - Oskar Jönsson

Contents

1

Introduction

In the introduction chapter an overall background is presented together with the purpose. Lastly, a light overview of the work process is displayed.

1.1 Background

Axis Communications AB created the world's first network camera back in 1996, [Communications, 2016a]. Since then the surveillance competition has thickened and still Axis strives to grow even further within the coming years. The growth is depended on the innovation within the company.

As a part of the innovation and growth, object detection with RAdio Detection And Ranging (RADAR) technology is being explored. The car industry is lowering the costs of radar components, making them a viable hardware to incorporate within the cameras. How can radar improve surveillance and create new business segments? That question has led to a number of new master thesis projects, along with internal teams researching this new field. One challenge is to graphically visualize the information. The aim of this master thesis is to find a way to visualize the result of object detection, combining radar technology with the image based object detection technology used today. The design approach will have a user experience focus.

Object detection today

Today the surveillance camera system uses an algorithm based on pixel changes for object detection. The changes in the pixels of the camera stream trigger an alarm. The alarm is then accounted for by an operator. The algorithm system is called Motion Object Tracking Engine (MOTE) and is an internally developed system for detecting and tracking objects in real time. It can divide the environment into foreground and background. It can detect changes and pinpoint objects within a scene. MOTE provides data about the changes, it does not visualize data.

Since the algorithm is based on changes in the scenery, a passing shadow might trigger an alarm, a false alarm. The false alarms cause problems, this in the sense that an operators focus is drawn to a false alarm. This also poses a security risk, since an operator might not take the alarm seriously in the case of a real trigger. As an attempt to reduce false triggers a radar is incorporated in the camera system.

RADAR combined motion detection

The radar (RAdio Detection And Ranging) is a classic *object detection* system that sends out radio waves that then bounce back and can be measured. The *radar* system gives a lot of valuable information, the rate of change of distance gives the speed of the object towards the sensor as well as the distance. By measuring with a high frequency the radar can determinate the direction that an object came from. [Budge and German, 2015]

The radar can complement the data gathered from *MOTE* and together make more accurate decisions. The dual system object detection can be based on both the image and the radar information. Theoretically, this will limit the number of false alarms, and create a more accurate object detection. With the new combined data, false alarms might be reduced. To be able to visualize the object detection in a new and user intuitive way a design process has to be conducted in order to understand the needs of the *end user* and what information to present.

The design process was based upon the new combined system with a *user-centered approach*.

1.2 Purpose

The purpose of this thesis has been to graphically visualize the combined information that the radar and camera provide by applying a *user-centered design* (UCD), process.

Also, this thesis evaluated the possibility to follow a user-centered design process with limited access to the end user. The end user was only involved in the validation of the final *prototype*. The questions this thesis was based on were the following:

- How can the information be graphically visualized to fit the demand of the end user?
- Can a user-centered design process be used with limited access to the end user?
- Will the iterative prototype phases improve the design process?

Literature study

A literature study was conducted to find related work and relevant methods for the thesis. Links and information to the connected work can be found in the bibliography.

Design process

The iterative design process was separated into five phases and can be seen in figure 1.1 below.

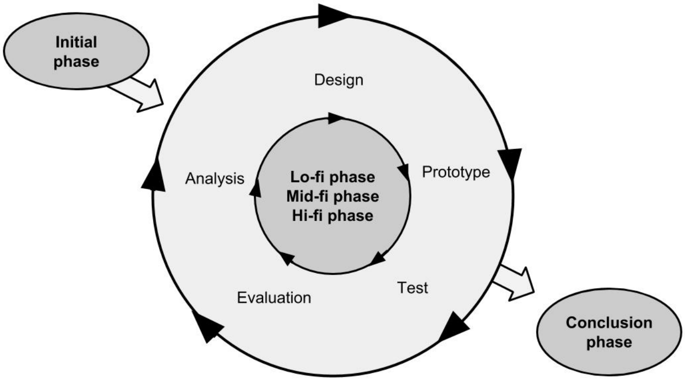

Figure 1.1 The iterative design process, The Initial phase and Conclusion phase are outside of the iterative cycle. The inner circle phases are iteratively performed and labeled accordingly.

Initial phase The first phase focused on defining the end user. This was done by interviews at the company and studying internal data. The result created a picture of who the end user was. Based on the information about the end user a *workshop* was conducted, resulting in a prioritized hierarchy of important features to include in the design. The phase concluded what *features* the low-fidelity (lo-fi) phase should contain.

Low-fidelity phase The second phase consisted of creating lo-fi paper prototypes based on the features from the initial phase. They were then tested and evaluated resulting in a rough feature design layout.

Mid-fidelity phase The third phase consisted of creating mid-fidelity (mid-fi) prototypes using a prototyping tool. The mid-fi prototypes were based on the results gathered from the lo-fi phase. The mid-fi prototypes were then tested and evaluated. This phase resulted in one digitalized design of the *end prototype*.

High-fidelity phase The fourth phase consisted of implementing the high-fidelity (hi-fi) prototype based on the mid-fi result. This prototype was then tested on the end user, evaluated and altered. This phase resulted in an end prototype, similar to a real product.

Conclusion phase The fifth and final phase consisted of evaluating the results gathered throughout the design process and generating the conclusion of this thesis. This phase is reflected in chapter 8, Discussion and 9, Conclusion.

1.3 Thesis delimitation

The limitations and delimitations are concluded below.

- The input to the system are a fusion between radar and MOTE. The input is a simplex communication channel. The graphical presentation of information is intended to be applied in the camera software and therefore no user interaction is possible. The thesis will only include visualizing the input information, not interacting with it.
- Object detection algorithms are not included in the thesis since they are out of scope.
- The product is intended to be used solely outside, hence this thesis will not cover any indoor use cases.
- The demo created by this thesis will be for one camera view only and the thesis will not include further studying of scalability of the design concept.
- The thesis will not cover the area of installation of the product.
- Due to company related circumstances the end user will only be involved in the hi-fi validation process.

1.4 Related work

This thesis is a continuation of a master thesis by *Madeleine Boström* and *Tobias Claesson* also conducted at the same company. [Boström and Claesson, 2017] Their master thesis involved processing the radar input together with MOTE to create a fusion of the two to *reduce false triggers*.

1.5 Distribution of work

The thesis has been conducted by both of us, working in parallel. The work has been divided at times, but evenly spread between the authors. In general, we have worked divided at the same task, only to summarize and create an efficient workflow.

2

Theory

The theory is presented below. The literature study conducted created the base for the thesis and concerned both the user, the measuring of data and principles for visualizing the result.

2.1 User-centered design

The term UCD originated from Norman and Draper. Their main point was the importance of having a good understanding of the user. This does not necessarily mean actively involving the user throughout the process, [Norman, 1986]. The user can be more or less involved in the process depending on accessibility and type of project.

UCD can be applied to all types of design processes as long as there is an end user, someone who actively uses the product. UCD works well with agile methods since the user may change their mind during the process. There are two main benefits of using UCD compared to linear processes. The amount of guess-working is minimized and the probability of developing a product that meets the need of the user increases when an active focus lies on the user. UCD is widespread and a commonly used process for developing with a user focus. Preferably the process involves the end user with the development team, defining demands and participating in the design process. UCD focuses less on formal methods of gathering requirements and specifications than other project based processes. In other words, the UCD design process can be described as a more flexible iterative methodology than a normal linear design process. [Kumana and Dickinson, 2014]

2.2 User

According to Eason and supported by Abras, the user can be divided into three groups; primary, secondary and tertiary. The primary user are the user that will physically use the product. This could be the operators using the product to monitor areas. The secondary user are the ones that use the product through an intermediary. This could be the guards on the field having contact with the operator. The tertiary user are user that will be affected by the use of the product or the user deciding to buy the product or not. [Eason, 1987], [Abras et al., 2004]

The needs, demands and wants of the user, guides the UCD process. It is also of high important that the user can easily and intuitively use the system. Furthermore, the user's demands and limitations have to be considered throughout the process. Despite the extent of user involvement in the process. [Ritter et al., 2014]

Methods for UCD without access to the user

As mention earlier UCD does not have to involve the user directly. Opposed to methods involving the user actively there are also methods that do not involve the user actively. Such methods could be *proxy interviews*, personas, environmental studies, and *activity/task analysis*. These methods can have a deep impact on the design outcome, together with design guidelines and norms such as *General principles of graphical visualization*, *Color principles*, *Gestalt principles* and *Norman's design principles*. These guidelines and norms are presented later in this chapter and they form a base for the final design. The most important thing is that the user must, in one way or another, be in the center of the design process at all times. [Abras et al., 2004]

Proxy interviews A method of interviewing. The definition used in this thesis is, *"An interview with someone (e.g., parent, spouse) other than the person about whom information is being sought."*,["Alcser et al., 2017]. The method can be used when a direct interview with the sought person is impossible. A proxy interview can be conducted with a person that has knowledge about the sought person and some information that the sought person has. The key to unlocking the information sought after is to find the most relevant person and have prepared questions. The aim is to find the sought information, e.g. in this thesis it was to identify the end user through the knowledge of the interviewee. A dynamic interview approach is suggested to be able to adapt to the given information and identify the end user. ["Alcser et al., 2017]

Task analysis Using an activity/task analysis to involve the user can be done with or without an active involvement of the user. The activity/task analysis can be used as a way to get to know the user and getting an understanding of the user's needs by analyzing what tasks the user has and how the user works with them. [Ritter et al., 2014]

2.3 User experience and usability

The relationship between *User Experience* (UX) and the *usability* of a product is a complex matter, the ISO definitions says:

User experience *"A person's perceptions and responses that result from the use or anticipated use of a product, system or service."*, [Standardization, 2009].

Usability *"The extent to which a product can be used by specified user to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use."*, [Standardization, 1998].

User experience aims to keep the user subjectively positive. The operator should feel content using the system, this calls for easy interpretations of the graphical visualizations. The scarcity of interaction reduces measurement of usability aspects and creates a greater need for measurement of the user experience. User experience and usability goes hand in hand and overlaps since they both strive for the user's satisfaction of using a product or a system.

Usability strives to ensure the ease of usage and interaction with a product or a system. As an example, an operator can use a joystick to tilt the camera, or a slider function to zoom in or out. This thesis will not include user interaction with the prototype. Therefore usability in the sense of interaction will not be in focus. Usability in the sense of using the product without interaction will be a part of this thesis. The product will result in a video stream the user can watch.

2.4 Measuring user experience and usability

It is important to define a way to measure both UX and usability. When working with a product it can be desirable to be able to present a numeric value of improvement e.g. a new product could be *70%* faster than the older version or e.g. twice as reliable. When working with a design it is not always simple to produce these desirable numbers and at times other ways of measuring might have to be taken. In this section performance metrics and subjective data collection will be explained to give a clear view of how the prototype was measured.

Subjective data collection

Subjective data aims for data that is hard to put a number on. It is in a sense soft information such as feelings and first impressions. Three types of subjective data collection methods are presented below, namely *think-aloud*, *Discussion and debriefing interviews*.

Think-aloud Think-aloud method is a time consuming subjective test method that can be used to gather subjective data from test subjects. Jacob Nielsen wrote, over 23 years ago in his book "*Usability Engineering*", [Nielsen, 2012], "*Thinking aloud may be the single most valuable usability engineering method*", and it still applies today. This is because the human behavior changes much more slowly than the technology and therefore the approaches to studying the human behavior remain the same or slightly modified.

While performing the think-aloud method, the test subjects have to mention what crosses their mind while performing a task. The observer will then get access to how they approach the given task and where the design accidentally confuses the test subjects. [Nielsen, 2012]

The benefits of this method are that it is cheap, easy to learn and it can also be easier for developers or designers to relate to where and how the test subjects encounter obstacles with the implementation or design. The downside with this method is that thinking aloud is an unnatural thing for the test subjects to do and therefore it can discourage the test subject. The test subjects subconscious might filter what they say in order to avoid appearing less clever. The observer might influence the test subjects by disrupting them to clarify or ask leading questions.

With all of this in mind, the think-aloud method still stays number one usability tool for the designers and developers today. [Nielsen, 2012]

The think-aloud method was used during the lo-fi and mid-fi prototyping phase.

Discussion and debriefing interviews Discussion and debriefing interviews is another subjective way of collecting data from test subjects. This differs from the think-aloud method in some ways. The think-aloud method is more instant and displays the user's thoughts directly during the performance of a task, without the involvement of the test leader. Whereas the discussion and debriefing takes place after the completion of a task and may abbreviate and focus on certain subtasks and decisions. This method was used in all prototyping phases.

Test plan

Measuring the UX and usability can be facilitated by following a *test plan*. To cover the full spectrum of the user experience the test plan should consist of both performance metrics, raw numbers to be measured, and subjective data collection, personal perceptions. Benefits with using a test plan are the structural performance, eliminating the risks of forgetting questions or tasks. It is also a good way to present the set up for the test when presenting it to a customer.
The parts of a test plan can be seen below.

Purpose, a short presentation of the product and the overall benefits with the test.

Background, a concise text consisting of the most important background that the test participants need to know.

Tests questions, A list of the test aspects or questions that has to be tested or answered.

concluded questions, A list of the test aspects or questions that has to be tested or answered.

Data to collect, Describing information about how and what information the test strives to collected to be able to find an answer to the test questions and fulfill the purpose.

Test details, For instance a matrix displaying valuable information about the test. This can include the different steps of the test, the questions mapped to the steps and preferably a time approximation for each step.

Test, Another matrix with an overview of the test. Information about what is to be done when and what material that is needed for the different part of the test.

Participants, A short description of the target test group and who is participating in the test.

Environment/tools, A list of the materials needed for the test. Information about where the test ideally should be performed etc.

Test roles, Describes what roles are needed for the test and what responsibilities they have during the test.

The test plan can be used as a base for e.g a *validation testing* procedure. Together with *Discussions and debriefing interviews* a test plan can cover the extensive testing needed to be able to validate a product or prototype.[Tullis and Albert, 2008]

2.5 Workshop method

To be able to narrow down the tasks, a feature analysis workshop method was used, namely, the KJ methodology.

Methodology of Jiro Kawakita

The KJ method, named after its inventor Jiro Kawakita, sometimes referred to as *affinity diagram*, is often used to derive priorities when resources are limited. As in a design process for example. The time resource is often limited and it is highly favorable to prioritize where to focus and direct efforts accordingly. According to Jared M. Spool, [Spool, 2004], an American expert at usability and design, there are eight steps that can be used on any test group within any field of work.

Determine a Focus Question, Every session has one focus question. If you need to answer several of them divide them into separate sessions. The focus questions could be; *Who are our user? What features do we need?* or *How do we present information?*

Organize the test group, Gather people with different perspectives. Estimate an hour per session.

Use Sticky Notes, Ask the participants to write down as many items they can think of, on sticky notes, connected to the focus question.

Use the wall, When all participants have finished putting sticky notes randomly on a wall. Then let the participants read each other's contributions. The participants can at any point add a sticky note to the wall.

Sort into target groups (TGs), Ask the Participants to sort the sticky notes without any discussion about neither the sticky notes nor the TGs. The participants can move a sticky note into whichever TG they see fit.

Naming the TGs, The participants individually names, not sentences, each TG.

Voting, Repeat the focus question so the participants are reminded to what they vote for. Then the participants individually write down the top three most important TGs and choose the best name that represents the group. Then the participant will rank their three TG names from most important to least important. The participants will take turns to mark their choices on the TG name sticky notes.

Ranking the most important TGs, All sticky notes within one TGs receive the number of points the TGs got in the voting. Now the sticky notes are placed in a hierarchy according to their points. The sticky notes are put in front of the participants and they can now suggest that two sticky notes merge. All participants must agree for them to merge and they need to be identical, subgroups does not qualify. One of the sticky notes are removed and its score is added to the remaining sticky note which rises in the hierarchy. In the end, there will be fewer sticky notes ranked in a final hierarchy that all participants agree upon.

The important and effective part of the process is saving the discussion until the end of the session. This eradicates discussions about the least important issues. This was included in the KJ method that was used in this thesis. [Ulrich, 2003]

2.6 Prototyping

Up until recently the prime method of explaining the design to customers was through documentation. This meant that with a complex interactive structure these documentations could be around 100 pages. A cognitively tiresome way to perceive information. Due to the process of creating these document, minor changes or sometimes even vital information could be unnoticed. And as a result, the document could quickly become outdated. The work put into these documentations quickly escalated and made it hard to motivate the benefits of documenting the design due to the changing phase of the design in early stages. [Adiseshiah, 2016]

Today prototypes are replacing the use of such design documents. It is hard to compete with the rapid development in prototyping and the iterative design process it encourages. To be able to show the design to the customer as well as to the development team offers a huge advantage. It is easier for people to grasp a concept if it is visualized rather than a documentation. [Adiseshiah, 2016]

An approach to prototyping is to separate into three stages. The stages have different focus areas and are divided into low-fidelity prototypes, mid-fidelity prototypes, and high-fidelity prototypes. Fidelity refers to the level of detail and functionality the prototype incorporates. [Pacheco, 2014]

A common method to combine with the design process earlier phases is the *Simplified user testing* and the *Narrowed-down prototypes* that can be found in *Discount Usability*. These are presented below.

Discount usability

20 years ago, Jakob Nielsen, within the field of user testing and agile development, developed a way to perform the simplified user testing and narrowed-down prototypes, [Nielsen, 2009], a cheap method that gave a great result for a project that did not have a huge budget for testing designs. The discount usability contains three important aspects:

Simplified user testing Simplified user testing focuses on qualitative data with a low amount of participants in the test group. About 5 test subjects are thought to be optimal. Simple user testing also promotes usage of the thinking-aloud method where the test subject speaks openly about the internal decision-making while performing tests on a prototype. This method was used throughout every design phase. [Nielsen, 2012]

Narrowed-down prototypes Narrowed-down prototypes refer to fast creation of different low and middle-fidelity prototypes with short development iterations. They should be easy and quick to put together and through testing, they strengthen the full user experience of the final product. This method was applied during the initial creation of lo-fi and mid-fi phases.

Low-fidelity prototype

Lo-fi prototypes are commonly used as the first visual step in a new design. They are simple, and as such, they are often cheap, easy and vague. The test subjects will not get hung up on details, such as colors or fonts.

A lo-fi prototype normally consists of papers and sticky notes. This ensures that realistic user interactions will not be visualized. Instead, the idea of the design will be sketched and easily improved. It is cheap to create lo-fi prototypes and a lot easier to spot design flaws, and improve the product without being overly time consuming or costly. It is through the lo-fi prototype *iterations*, a decision about what and how things should be incorporated is determined. A quick and easy step towards a conceptual design without to much focus on the visual details of the layout, text size and fonts etc. The important part is to get a basic idea of what to evolve in the next iteration. [Farrell, 2015] An interesting method that may be applied to lo-fi is the *Exploratory testing*.

Exploratory testing Exploratory testing aims to compare different prototypes to each other. The test method itself can be used at all stages of the process. When used in the early stage, exploratory testing aids in preventing making wrong decisions early in the design process. Exploratory testing can be used to measure the *cognitive visualization* of a single element of the design to see which version is preferred by the test participants. [Rubin and Chisnell, 2008]

Mid-fidelity prototype

Jerry Cao, co-author of the book *The Guide to Mockups* describes the different stages of fidelity prototypes as a range from lo-fi to hi-fi. To give a prototype too much fidelity too fast might leave avenues unexplored due to the cost of changing a hi-fi prototype. Cao explains the mid range, mid-fi prototyping phase, as *"being lazy in a good way"* or *"responsibly lazy"*.

Exclude the lo-fi sketching prototypes and consider the digital prototypes. There are according to Cao two types of these. The hi-fi prototype that should represent the end design exactly and the mid-fi prototype that can be described as a transitional mock-up phase and hence should not take too much time to create. [Cao, 2015]

A mid-fi prototype is more precise than a lo-fi and more flexible than a hi-fi. It visualizes the design concept in a more styled, less conceptual manner. Using mid-fi prototypes can reduce the risk of investing a lot of time and money in creating components that will not be used in the final design by detecting the design flaws in an earlier stage. A drawback, especially in UCD, is that the end user and clients might not be as impressed or convinced by a mid-fi prototype as by a hi-fi where it looks like the final design. [Cao, 2015]

The mid-fi prototype should be a digitized version of your sketches. It should not contain much functionality and could be implemented in a prototype visualization tool, image processing tool or a similar tool for digital prototyping. The mid-fi prototype builds on the conceptual ideas extracted from the lo-fi prototypes. [Pacheco, 2014]

Comparison testing Comparison testing is not used for one specific stage of a product development but can rather be used in combination with other types of tests such as exploratory testing. The comparison test is used to compare two or more prototypes. Comparing two prototypes in a later state of the project gives inputs about the design while comparison testing in earlier stages rather gives a general direction of the project. [Rubin and Chisnell, 2008]

High-fidelity prototype

At the hi-fi prototype stage, it is all about the final design such as text fonts, sizes, dimensions, and color schemes. Cao calls this stage of fidelity *"pixel-perfect"* since it is about the actual pixel design at this stage.

Some of the benefits of hi-fi prototyping are that it is easier for non designers, end user or customers to really get a feel for what the final product will look like. It can be hard for people without knowledge about what fidelity means to separate a mid-fi and a hi-fi prototype since both are digitally implemented. When using hi-fi prototypes the developer implementing the final product can often use the mid-fi prototype as a specification. Thus solving the problem about describing design in a document. On the downside, it is time-consuming and costly to develop a hi-fi prototype so you should make sure that the previous steps are well executed. What would have taken minutes to change at a lower fidelity stage could now take days. [Cao, 2015]

Validation testing Validation testing or more precise, user acceptance testing is used to determine to what extend the prototype fulfills the user's needs and demands. A good looking prototype might still fail the acceptance test if demands have been misinterpreted or down prioritized. So following the careful steps of an *UCD* approach with research should, in theory, produce a satisfied user and a confirmation from the *validation test*. The validation testing can be included within a test plan to gain structure. [Tullis and Albert, 2008]

2.7 Graphical visualization and design principles

Graphical visualization relates to presenting data or information to the user. In this chapter, several ways of doing so are presented. *Infographics* and the concept of the *Useful Field Of View* (UFOV) will be explained. Furthermore, color, gestalt and design principles that were guiding the design process for this master thesis will be presented.

Infographics

Icons and pictures are two examples of infographics, graphical ways to present information. Ranging from the cave paintings to the smileys of today, *" the infographic uses visual cues to communicate information."*, [Lankow et al., 2012]. Another description of *Information graphic*, known as infographics is stated by Mark Smiciklas as *"visualization of data or ideas that try to convey complex information to an audience in a manner that can be quickly consumed and easily understood."*, [Smickilas, 2012].

Cognitively, the brain perceives each letter or word as a symbol. When reading a text the brain adds several of these symbols together to get the full context in a linear manner. When the brain processes an image it processes the data in parallel. Meaning that only one impression of the image is perceived and processed.

Therefore if several text symbols could be replaced with one infographic or icon the brain would be less strained which would improve the user experience of using the design. When the brain processes an image all data processed at once instead of a linear manner as with text, [Smickilas, 2012].

As a simple example: an abstract image of an analog clock showing the time 11:43, see figure 2.1. Perceiving the time from the image is done with a simultaneous process and takes no time.

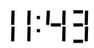

Figure 2.1 An infographic displaying a digital clock as opposed to reading *Forty three minutes past eleven o'clock*

If you instead had to read, *forty three minutes past eleven o'clock*, you would have no trouble to extract the data but the process would be sequential and would take considerably longer time, [Smickilas, 2012]. Highly important to consider when designing graphical visualizations for a surveillance system where time often is of the essence.

Researchers show that generally, people do not like to read of a screen compared to analog books. Jacob Nielsen established that the average person will read no more than 20% of the words on a web page. This could be an indication that using a lot of text to present information online might not be the best way if you want the attention of the reader. [Smickilas, 2012]

Useful field of view

UFOV is an area within the *Field Of View* (FOV). FOV is the total area that can be seen at a given time, including the periphery. UFOV is the area where the vision is most detailed.

The definition of UFOV is the area where the human eye can rapidly perceive information without having to use head or eye movement. UFOV is often presented in how large the area is expressed in degrees. The size of UFOV is personal and can adjust depending on the tenaciousness of the tasks.

Stressful conditions and/or constant concentration might create a phenomenon known as *tunnel vision*. *Tunnel vision* narrows the UFOV, excluding even more information. In 1985 a study by Williams [Williams, 1985], compared the performance in detection objects in the periphery while performing two tasks. The study concluded that if a person is performing a task requiring intense concentration, the detection rate was 36%. In comparison to an easy task, low cognitive load, the detection rate was 75%. [Ware, 2004]

Peterson and Dugas performed a study in 1972 [Peterson and Dugas, 1972] that suggested that a moving target catches the eyes attention and shifts the UFOV easier than a static target. This finding is highly used in systems today, as an example moving advertisements or twinkling lights, a classic interface feature for attracting attention. [Ware, 2004]

General principles of graphical visualization

When it comes to graphical visualization there are some general rules described by Tullis and Albert in their book *Measuring the user experience*. [Tullis and Albert, 2008]

- Using numbers are generally good for presenting information. However, the precision should reflect the reality. For example 20.*00* seconds or 100.*0%* are almost never appropriate. In most cases, less precise numbers are both clearer and more relevant for the user.
- Colors should not be used alone to convey information. Colors should be used together with positional information, labels or other cues to enable someone who can not distinguish color to interpret information.
- Labels should be displayed horizontally when possible since vertical labels increase the user's mental and or physical workload.
- Do not overload the information presented. Just because it is possible to compress all information into one graph or screen it does not mean you should.
- Using text on a background with visual noise can make the text very hard to read and it is therefore not recommended to use text with a changing background. [Johnson, 2010]

Color principles

In the book *Designing with the mind in mind* by Jeff Johnson, [Johnson, 2010], there are several guidelines concerning colors.

- Distinguish colors by saturation and brightness as well as hue, do not use subtle coloring. Contrast is important when choosing colors, a simple greyscale will give an indication if the contrast is proper.
- Use color pairs that color-blind people can distinguish, for example, colorblindness can make it hard to distinguish blue from purple, red from green and green from the khaki. To ensure good contrasts use a dark color complemented with a light colored background or vice versa.

- Use color redundantly with other cues, do not use colors as the only indication of marking something. The color is a great supplement but should generally be sparsely used on its own.
- Use distinctive colors, there are three opponent colors pairs, red versus green, blue versus yellow and black versus white. It is easier for people to distinguish colors that are among one of the opponent colors. If a color is a mixed color the color-perception will have to work harder and therefore it is harder to distinguish the color. Opponent colors are also mentioned as complementary colors, pairs that create well contrasted if greyscaled.
- Separate strong opponent colors, the usage of opponent colors creates a shimmering sensation which often is disturbing for the user. This should, therefore, be avoided.
- Red, yellow and green, red is a standard color for error, it can be seen in for example error messages or warnings. Yellow is often used as a transient warning color and can, for example, be found in the traffic lights as the standby state. Green is a general color of success. [Mulholland, 2016]

It is important to remember that there are several aspects that the designer needs to keep in mind when designing for user. Here are some of the example of aspects Johnson mentions as affecting the perceived color. [Johnson, 2010]

- Pale colors are harder to tell apart than bright colors.
- Size affects the ability to tell colors apart. It is harder to perceive the color of small or thin objects than on larger patches.
- Separation of objects affects the ability to separate the colors, especially if there is a need for eye movement for the user to see both objects.
- Computer screens may display colors differently and therefore what are perceived as a distinct color on one screen might be perceived totally different on another screen.

Gestalt principles

Gestalt is the German word for figure and the seven gestalt principles aim to explain how human visual perception works. Today they are thought of as a descriptive framework within the field of visual perception. There are two principles that concern the visual grouping of objects. [Johnson, 2010]

Proximity, the distance between objects affects our perception about how the objects are organized. If objects appear near each other they are perceived to belong to the same group or as rows instead of columns, see the top left part of figure 2.2.

Similarity affects our perception of group belonging. This since objects that look similar appears to be part of the same group. This can be applied to everything from designing a form to deciding the usage of fonts for text. An example is distinguishing triangles as a group even if they are not close to each other as can be seen in the top middle part of figure 2.2.

As a complement to the grouping principles, there are also four principles concerning perceiving objects and structuring the received data. [Johnson, 2010]

Continuity describes the visual tendency to resolve ambiguity and fill in missing data to create a whole object. Even if only parts of the object is visible it can be perceived as complete. This could be displaying as two curved line crossings. They are not perceived as two lines making a 90 degree turn, see top right part of figure 2.2.

Closure, the visual system automatically tries to close figures that are not complete. This could be drawing just the corners of a shape and it would still be perceived as a complete shape as can be seen in the bottom left part of figure 2.2.

Symmetry, the ability to parse complex scenarios so that it reduces the complexity. Usually, the data in the FOV can be interpreted in several ways, the vision organizes and interprets the data to simplify it and obtain symmetry. A typical example here is two squares overlap each other slightly but still looks like two squares and not like three squares as can be seen in the bottom middle part of figure 2.2.

Figure/Ground, our mind structures the received data into a *Figure*, the foreground, and a *Ground*, the background. The foreground contains all elements that are the focus of the attention and the rest becomes part of the background. The characteristics of the scene also influence the parsing in the way that if a small object overlays a larger one there is a tendency of perceiving the smaller object as the figure and the larger as the background. This however also depends on where the viewer focuses. An example can be seen in the bottom right corner of figure 2.2.

As a contrast to the six principles mentioned above relating only to static figures the last principle concerns moving objects.

Common fate is a principle related to the proximity and similarity principles about grouping, but in the common fate, the grouping is related to moving objects. If only some of several identical objects are moving they are perceived as a group.

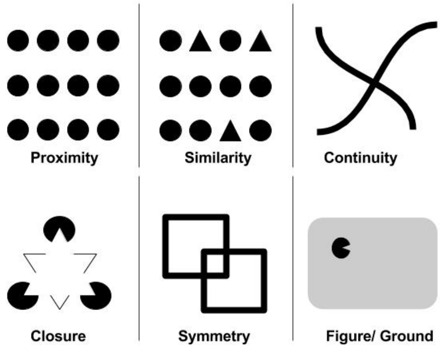

Figure 2.2 Top left (Proximity), middle (Similarity), right (Continuity)
Bottom: left (Closure), middle (Symmetry), right (Figure/Ground)

These gestalt principles are almost always used in combination. When using all of them at once there can be unintended visual relationships implied by the design. To avoid these viewing the end design from the different principles perspective separately can be an efficient way to detect unintended relations. [Johnson, 2010]

Norman's design principles

A classical way to conceptualize usability is Donald Normans design principles, [Norman, 1986]. Since usability is a part of the user experience. Therefore the principles were followed in the design process.

Visibility, functions that are visible will more likely be used correctly in comparison to "out of sight" functions.

Feedback, any form of feedback on the performance is valuable for a user, tactile, verbal, audio, or a combination.

Constraints, restrictions enables the user to interact in the right way at any time.

Mapping, relationships between control and effect.

Consistency, it is important to use similar elements for similar tasks. Consistency can be achieved by following graphical guidelines.

Affordance does the element introduce the user to the usage, invites interaction because the outcome is known.

3

Method

The approach performed in the master thesis is presented below. The methods and theory used throughout the process are presented in chapter 2, Theory.

3.1 Approach

This thesis explored the use of UCD with limited access to the end user. It was not possible to involve the end user before validating the hi-fi prototype, due to company regulations. A combination of UCD methods and user definition methods in an agile design process were used.

The absence of an end user leads to a problematic definition, the end user has to be devised based on second-hand information. The second-hand information was derived from proxy interviews and the proxy interviews created the user-centered background needed for arranging a workshop. The workshop generated a hierarchy of desired features. The feature list was narrowed down to features related to information visualization.

The narrowed down features were then visualized and presented in a low-fidelity test session. The feature design test session resulting in a couple of lo-fi prototypes. The best features from this phase were passed on to the mid-fidelity prototyping phase. The decision to use a mid-fi prototype instead of going from low to high-fidelity directly was due to the lack of access to the end user. It was easier and cheaper to make drastic changes to a mid-fi prototype rather than to a hi-fi prototype. The mid-fi prototypes were made in a prototype visualization tool and after testing and evaluating these prototypes one hi-fi prototype was created by combining the best parts of the mid-fi prototypes.

The hi-fi prototype was tested on the end user and evaluated. Presenting information for the discussion and the conclusion of this thesis.

4

Initial phase

In the initial phase, the end user was *identified, defined and concluded*. Features connected to the defined end user was prioritized by the *KJ method workshop*. The aim of this thesis only involves the end user at the final part of the process, validating the result.

4.1 Identifying the end user

To be able to work in a UCD process, the end user had to be *identified*. Due to the *sales model* the manufacturer is not in direct contact with the end user. Some employees have knowledge about the user, but the general picture of the user is vague. Some departments work with a specific user, another department knows another one. To be able to identify who the end user for this thesis would be, a couple of inquiries and processes were executed, these are all presented later in this chapter.

User-centered design constraints

Important company related factors that affected the UCD process was the indirect sales model and *the system structure*.

The indirect sales model Partners retail and distribute cameras and surveillance software to *system integrator* and *resellers*, not directly to the user. The system integrators and resellers sell a full-service system to the customer who usually employs the user. The customer might be a large retail store, the user could then be an *alarm operator*. The indirect sales model ensures the innovation within the company and also the professionalism and competence of the specialist in every stage of the model, from production to end customer. [Communications, 2016b]

This means that the production has little or no contact with the end customer; and more importantly for this thesis, the end user. The lack of communication between the developer and the end user leads to problems when developing and innovating new products with a UCD approach.

The hardware and software system solutions A broad variety of hardware and software system solutions are provided to retailers. The customers have different needs and the systems are supplying a wide customer range with solutions and in extension an even wider range of users.

As an example, System 1 and System 2 aims to satisfy different customer segments. System 1, avails installation of several cameras at once. The software is used by a system integrator, small business. System 2, is for medium and enterprise size installations, larger businesses. The larger installation requires multiple system integrators and further needs an operator for day to day use of the system. These systems aim to satisfy a different customer segment and as a result, it also targets different users.

For installing cameras hardware installers and software integrators are needed. All the systems are adapted for different sizes of companies. As an example, a store owner needs to feel that they can choose a suitable solution for their store. For larger installations, multiple system integrators and operators are required. For some installations, *security guards* are key personnel, important to enhance the security.

As a whole, all of the system solution systems aim to different customer segments. And by extension, also targets different users, as defined above. The data was used firstly, to present the different users, and secondly to define the end user. The possible users are; system integrators, hardware installers, resellers, security guards, and different kinds of operators.

User-Centered Design at Axis As a response to the problems that might rise with distance to the end user and the many different types of users, there has been an intensified effort on designing for the user. As a part of the procedure, a whole new department has been created with the sole purpose to integrate the user, by focusing on Usability, UX, and UCD. This project has been ongoing for about a year and is constantly growing. To collect information about the user we read the internal sites for the different systems along with personnel interviews.

Proxy interviews

We performed three proxy interviews with the *UX lead*, *UX expert*, and an *object detection specialist*. These are the personnel who had been in contact with end users. The data gathered was the base of an feature workshop. That leads to a definition of what from here on will be referred to as the end user.

Information from interviews The first interview, UX lead, created a vast picture of the possible end user, operators. There are mainly two different kinds of operators; *Surveillance Operator* (SO) and *Alarm Operator* (AO).

Both operators are located in a camera surveillance room, monitoring multiple screens. Mainly interacting with the system on a computer.

The SO is familiar with the surroundings of the camera and is constantly watching, scrutinizing the cameras video stream.

The AO acts on object detection alarms, triggered by the system. AO might be monitoring many systems at the same time and is required to make quick and correct decisions based on the data presented on the screen. Another important issue is the fact that the AO is not necessarily familiar with the surroundings of the alarm resulting in a need to rapidly build a cognitive picture of the situation.

Even though the use cases differ between SO and AO there are certain similarities. As an example, both needs to identify what triggered an object detection, when and where it triggered. As a possible solution, the interviewed suggested graphical visualizations such as *maps*, and *history data*. Together with the possibility to turn it on/off, toggle function.

The second interview, the UX expert, was conducted over the phone with an expert on customer behavior. Some pointers and directions from this interview lead to pinpointing the end user even further. For an example, some system integrators are valuable users. Since the sales model can be supplying a full system, with installation and continuous surveillance and follow-up. This case creates a wide range of users, with a broad competence and deeper knowledge in surveillance.

The third interview, object detection specialist, focused a lot on forensics and research. The forensic and research work is usually conducted after an alarm has been triggered. Usually without a time pressure. The focus is instead finding the object that triggered the alarm. A detailed and precise system is more important than a fast detection service. The user is more interested in smart interactions, rather than a smart system. They want to steer and make rational decisions based on background information together with the camera data. Another feature that was highlighted was the ability to filter data. Based on the presented scenario some data might be more or less important.

4.2 Defining the end user

The end user was defined and its different characteristics displayed.

Different end user characteristics

The prototypes aim to cover a customer segment not yet clearly mapped. Since the customer segment is broad, the end user can only be constructed by the gathered qualitative data. To be able to create the qualified end user there are several constraints to take into account together with the data from the proxy interview.

- The intended customer segment target is medium to enterprise sized businesses. This means larger corporations, eliminating the low scale user: *store owners* and security guards.

- The components within the camera will be costly and as such, they will be assumed to be part of a special product line. The specific product line is mostly sold to medium and enterprise businesses. Resulting in a strengthened elimination of the low scale user: store owners and security guards.
- The object-detection system works best outside. Therefore the predicted use case is outdoor surveillance. Providing trigger detection at big and important areas. This also implies medium to enterprise sized businesses, again eliminating the low scale user: store owners and security guards.

The system integrator, system designers and hardware installers are not the ones that are looking at the streamed video when an alarm triggers. And as such, they are not the intended end user for the visualization of object detection and will therefore not be part of the end user for this thesis.

4.3 Concluded end user

Based on the information gathered from the design constraints, proxy interviews, and system structure, the end users is concluded into the primary users, from now on referred to as:

Alarm Operators Commonly situated in an alarm center environment. Working with a large set of cameras, located at multiple locations. AO is assumed to be unfamiliar with the area surrounding the cameras. This end user also works with forensics and history, thus might require interaction and filters. AO directs information to personnel on the ground making it important to direct correct information.

Surveillance Operators Commonly a stationed security guard, getting information from a computer screen scrutinizing camera feeds and can act before an alarm is triggered. Works with a specific set of cameras, it is assumed that the SO has good knowledge about the surrounding area and the camera setup. SO directs information to personnel on the ground making it important to direct correct information.

4.4 Feature workshop

A workshop was conducted to brainstorm visualization object detection features that could fit the needs of the end user and prioritize them to guide the design process forward. The focus question for the session was *"What features could the end user need?"* in the context of the graphical visualization prototypes that were to be developed. The workshop was based upon the KJ method, presented in the method section, with some additional alterations presented below.

- The number of votes the participants could use were increased from three to four since there were many *Target Groups* (TGs) to be ranked.
- A step was added after naming the TGs so that the participants could ask for a clarification about the sticky notes if needed. This opened up for the writer to explain what feature the sticky note was describing without opening up for discussion.

A *prioritized feature list* was created and would function as the main input for the lo-fi phase. The whole session was performed in silence, enabling the participants to focus, think, and eliminating unnecessary discussions.

Five participants with different background knowledge and input angles took part of the workshop. The participant's background were two other master thesis students at Axis from Lund University faculty of Technical engineering, within the Information- and Communication technology program, and two object detection experts and an UX expert also from Axis.

An introduction presented both the end user and the focus question. Information about attributes and possibilities for the object detection system was presented. After the introduction, participants were allowed to individually write down features on sticky notes. Ideas from "*detecting objects moving in the wrong direction*" to "*Geographical information such as object positioning by longitude and latitude, preferably mapped in google maps*" were written down. The features were randomly put up on the wall for all participants to overview. The participants read each other's contributions and were encouraged to add more.

During a complete silence, the participants sorted the sticky notes into target groups (TGs) and then individually the TGs were named. Each TG constellation had several names, one example could be seen in 4.1 where all the green notes were names. After contemplating the different names, the participants were given a chance to split and merge the TGs. When satisfied with the names for the TGs, an individual rating was concluded. Creating a hierarchy of TGs and the related sticky notes were awarded the TGs summarized points.

Figure 4.1 One of the TGs after the participants named the groups. The green stickers is the related TG names, bundled. The yellow stickers is the features and the red markings is the corresponding points.

After discussions and scoring, a result consisting of a feature hierarchy based on the end user was determined.

Figure 4.2 The priority list, the result after the workshop. Stickers with high points are placed above lower points.

Factors affecting the result Many of the features that came up during the workshop were based on interaction with the interface. These features were eliminated due to the inability to interact with the system. The workshop did not explain the delimitation since it was a hardware limitation and it was not desirable to influence the participants with further constraints. Had the participants been told the result might have differed. The focus was the user prioritized features, and this would be forwarded to the future works section.

It is hard to tell if excluding any features involving interaction would have altered the result but it is something to consider.

Summary and top prioritized features

The initial phase concluded a end user consisting of two types; SO and AO. Based on these a KJ method workshop was conducted to find the most important features for the future design.

The workshop resulted in a prioritized hierarchy of the features the group as a whole thought most important for the end user. The top four features were:

1. The top feature, (27 points) related to geographic information such as a map. Information mapped to movement such as positioning, *history* and lastly, camera placement overview.

2. The second top feature, (26 points) related to being able to filter angles or areas to ignore, for a time period, to avoid false alarms.

3. The third top feature, (20 points) related to highlighting the important areas or rather darken unimportant areas, known as a *masking* effect. And further information about marking objects.

4. The fourth top feature, (14 points) related to *marking* and classifying objects in the image and information about history.

Conclusion

After filtering out the features that focus on interaction, such as an interactive map, filter areas and masking. Features regarded as interactive are outside of the scope, since the camera view is in focus. They are still important results of the thesis, and interesting as future work. Interactive features will not be actively involved in the design process, only complementing the concept along the way.

The top prioritized features were concluded into three categories that would be the foundation for the rest of this thesis:

Marking, the ability to pinpoint and highlight the area where the object was detected.

Identifying, the information about what object was detected.

History, the geographical history of where the object was first detected and the movements from that time.

5

Lo-fi prototype phase

In this phase *lo-fi feature design prototypes* were created based on the features from the workshop in the *initial phase*. The design prototypes were then tested, based on the cognitive visualization of several test participants.

The *interactive features* from the workshop were removed since they do not match the scope for this thesis. The features left were used as the foundation enabling development after demand.

5.1 Identifying key stages of object detection

The lo-fi prototype should strictly graphically visualize features. If there was a background to the visualizations, then the background could affect how the test participants interpret the feature design. Therefore a decision was made to work with a blank background, resulting in a clearer focus of the visualized feature in the testing. The features were concluded into three key stages: After an object detection, an operator; first need to find the object in the picture, secondly identify the object and thirdly decide what has happened.
Three key stages of object detection:

Marking, where is the object that triggered the alarm located?

Identifying, what object triggered the alarm? human, non-human or unidentified.

History, where did the object appear first, what happened before the Operator found it, what might happen afterward?

Based on the key stages above, lo-fi prototypes in the form of, *feature design prototype cards* were created. The cards visualized the feature designs in different ways. A simple and clean design was used to challenge the imagination of the perceiver. All together over 50 cards were brainstormed initially based on own thoughts and drawn in a uniformed simple design, see figure 5.1 for some examples. The inspiration was gathered from other products and solutions within object detection such as global positioning system (GPS) or games.

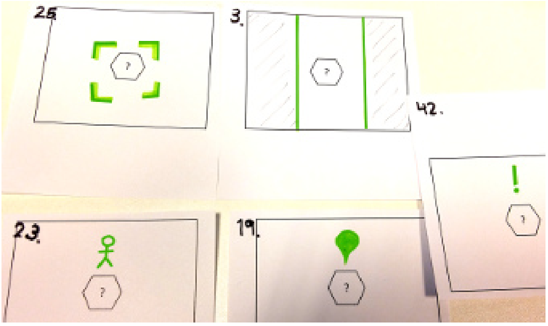

Figure 5.1 Some of the lo-fi feature design prototype cards. E.g. Number 25 in the top left corner is displaying an open box marking on the object. Number 23 in bottom left corner is displaying a human symbol.

The cognitive interpretations of the cards functioned as an exploratory test which resulted in the summarized lo-fi prototypes.

5.2 Exploratory testing of key aspects

The test was performed by 14 participants ranging from experts within object detection to master thesis students. The goal of the tests was to get the cognitive insight on the feature design cards. If theses cards were interpreted in a uniformed way by the participants, then the end user would probably interpret the information similarly.

The exploratory test session Exploratory testing in conducted by asking the participants how they interpret one feature design card and then comparing the card to others. An informal exploratory testing approach was used to be able to test the individual visualizations and create complete lo-fi prototypes by comparison. The *think aloud* method was used to further collect subjective data.

The test sessions were estimated for half an hour and were conducted with participants individually. The participants were greeted to a "Cognitive Workshop", the complete process involved four steps. An object had been detected in a camera view, the operator needs to know where and what had happened.

1. Marking the object, the participant analyzed every card in the category. The category consisted of different visualizations of distinguishing the object. The participant had to choose the top four most subjectively cognitive cards.
2. Identifying the object, again the participant analyzed cards. The current category was different ways to identify objects. The cards were mainly based on infographics marked directly over the object. Again the test subject had to

choose the top four. A discussion about the meaning of the infographics was initiated to classify the disregarded cards.

3. History, the third step included analyzing cards in the category of marking the history of the object. The test subject had to keep the top four on the table.
4. The last step was to combine the top marking-, identity- and history cards into the complete experience combination. The participant's combinations were the base of the lo-fi prototypes. This step forced the participants to elaborate around the cards with subjective data. Enabling them to form new perspectives and rediscover previously disregarded cards.

During the test, the participant was encouraged to draw their own ideas on new cards. Resulting in new cards that were passed on to the rest of the participants. This motivated the them to think more and collaborate with a greater content. The cards created were not always a part of the favorable collection.

5.3 Results and discussion

After ending the test sessions the total number of cards were approximately 80 since the participants added their own suggestions to the deck. This meant that a lot of data had been collected. To structure all the data a matrix of the top cards was created from each participant within *Marking*, *Identity* and *History*. Cards that resemble combinations of other cards were split into its subgroups. Example card number 51 was a combination of card number 3 and 32. Every card scored a point for each occurrence within participants top four cards in all categories. In this example, if one participant rated card 51 as top four both card 3 and card 32 would receive an indirect point each.

Marking

The marking cards were combined by similarity into groups, from *Alpha* to *Juliett*. For example, all shapes similar to a square were formed into one group, *Foxtrot*. The groups received the combination of the participating cards to indicate a rating. The different groups with their corresponding points are presented in table 5.1 below.

Table 5.1 The result after grouping the *marking* cards after similarity. The score reflects what group the participants preferred.

Group	**Content**	**Score**
Alpha	Marking the contour of the object.	8
Bravo	Underlines the object.	5
Charlie	Inserting a transparent mask of the surrounding left and right sides.	10
Delta	Direction indication below the object.	4
Echo	Different transparent masking of the surroundings.	14
Foxtrot	Square shapes.	12
Golf	Border indicator.	0
Hotel	Dyes the object.	6
India	Lines, locating the object.	4
Juliett	Highlights the surrounding area.	0

From this data all groups with less than 8 points were eliminated, leaving *Echo*, *Foxtrot*, *Charlie* and *Alpha* as the top four groups. Two of the groups, *Echo*, and *Foxtrot* contained transparent masking of the surroundings in combination with either marking an area around the object or the object itself.

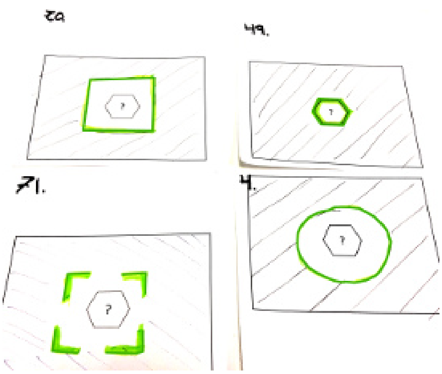

Figure 5.2 The cards within the *Echo* group; 70, 49, 71 and 4. They display the *masking* effect on different markings, e.g. number 49 in the top right corner is displaying a *contour* marking with a *masking* effect.

Echo contained cards where the surrounding area was transparently masked. It differentiates internally by marking the object in circles, squares or contours as can be seen in figure 5.2.

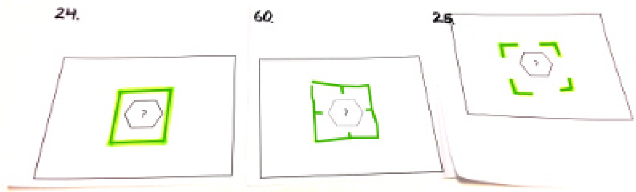

Figure 5.3 The cards within the *Foxtrot* group; 24, 60 and 25 are box shaped marking.

Foxtrot, figure 5.3, contained cards where the object was marked within a square shape, a box. This is a common method to highlight and visualize objects today by competitors.

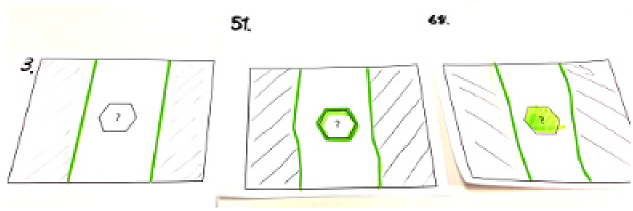

Figure 5.4 The cards within the *Charlie* group; 3, 51 and 68 has a vertical masking function.

The *Charlie* cards, figure 5.4, were characterized by the sides being vertically covered by transparent masking. Within the group, the object itself could be unmarked, contoured or dyed in combination with the transparent masking.

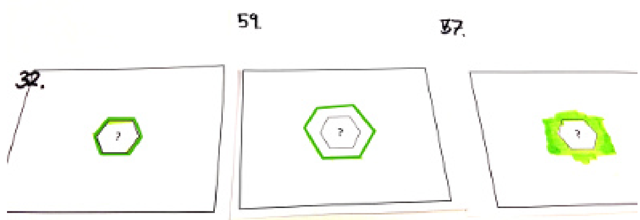

Figure 5.5 The cards within the *Alpha* group; 32, 59 and 57 has fitted or contoured markings.

Alpha, figure 5.5, contained cards where the object was fitted or spacially contoured.

Within *Echo*; among the masked cards, the card with fitted contours, 49, received 3 points whereas the circular marking, 4, received 10 points. This reflected what also was mentioned during the discussion that the participants wanted to see what happened around the object even if the rest was transparently masked. The circle was in many cases referred to as similar to the square options due to the intent rather than the shape. Regarding the masking, some participants voiced concerns about disregarding information since the software could be mistaken. Others thought it was a good way to find focus.

The cards grouped in *Foxtrot* were liked due to their clear box shape. The object was not covered by the shape and the possibility to scale was also expressed as a benefit. Two of the three cards in this group, 24 and 25, received approximately the same amount of point. The difference was a closed box compared to an open box. The third card, 60, however, received none. Participants commented on the visualizations strong similarities to a crosshair and that was not preferred.

The *Charlie* cards received the same amount of points. The discussion reflected that marking the object was preferred in the combination of masking the sides since it was clearer. Some participants wondered what the benefits of only masking the sides opposed to the *Echo* group were.

A fitted contour marking was highly preferred within the *Alpha* cards. The top card from this group was chosen to be part of several creations of new cards throughout the workshop part of the test session. This also meant it would receive an indirect point when one of the new cards received a point. This was the reason for receiving the highest amount of points within the group. Only one participant voted for this card on its own. The rest of the points were only meant for the card in combination with others. During the discussion, it was mentioned that using a thin contour as marking an object could easily blend into the background depending on what background was used.

Some participants pointed out that masking an area might be unnecessary, even unwelcome since information disregarded by the system might be overlooked by the operator. The masking could in a sense hide undetected suspects, preventing the operator from finding the suspects. The operator would want to make their own decisions based on the visual inspection, with support from the system.

From the participants perspective, the precision of the software was also pointed out to affect which marking could be most intuitive for the end user. By adding space between the object and the marking it allowed some deviations from the precision of the software without affecting the end user's experience of the product. Ideas to use *colors* to avoid the markings blending with the background came up during the discussions. By using complement colored markings would enable the end user to detect the target regardless of the color of the background. Almost all participants thought that all objects should be marked in the same way regardless of what type of object was detected. Some participants suggested that the use of color and size could be altered to separate differences of the targeted objects.

Identification

In the identity step, no points were given since the focus was on what the participant's associated with the identification symbols or how they responded to the usage of text. In regards to *human identifications*, the infographics were preferred over text or letters. The text variation was disregarded by, language issues, inelegant with many objects. It was also mentioned that if using text it should be used horizontally which is also strengthen by the theory section. But the text was also approved as being obvious, which the letters were not. Almost every participant mentioned that the letters were inexplicit and they were disregarded. This is also strengthened by the theory section stating that using text as an overlay on a background with visual noise can make the text hard to read. A camera stream falls under the category of the background containing visual noise. The symbols illustrating humanoids were evidently obvious, they received higher points. The symbol on card 47, see figure 5.6,was commented as a universal symbol for a human. It was also brought up that the human infographics could be used in combination with alert symbols or other data.

During discussions, the participants were unanimous in the decision that the infographic (**!**) meant interesting object. Most participants interpreted the symbol as a *warning sign*. A suggestion was that the object was detected in an *trigger zone*. As opposed to a (✓) when an object was in an allowed area.

The participants compared card 42 and 44 as can be seen in figure 5.6. They were thought to be similar even though 44 generally was preferred.

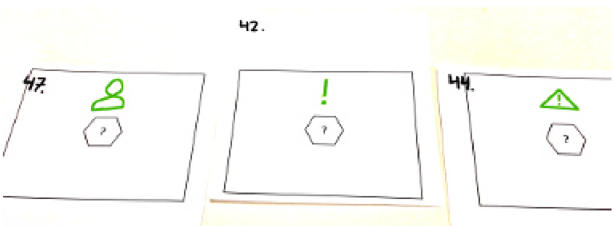

Figure 5.6 Card 47, 42 and 44 displaying symbols.

Most of the participants raised concerns that 42 would likely blend into the background and that 44 was better in the aspect of drawing attention to the object. The idea that 44 could be used together with colors and/or size to create an alert scale of what action to take as an operator. One participant voiced that 44 was interpreted as a higher alert than 42.

Some participants also agreed that 44 could be combined with human symbols, detection type.

Similarly another infographic (**?**) were established to signify an unknown or unidentified object. The cards 13 and 54 shows the same infographic but in different ways, as can be seen in figure 5.7.

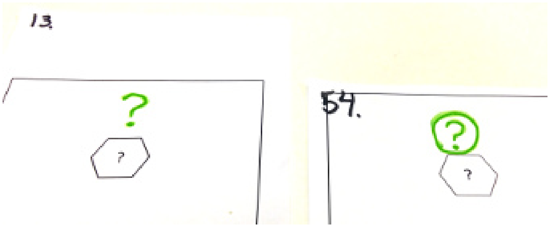

Figure 5.7 Card 13 and 54 shows the (**?**) symbol in different design infographics.

The same concern about the symbol blending with the background was raised and overall the participants preferred 54 since it draws more attention than 13. Since the participants did not feel that the object was dangerous when classified with 13 or 54 color and size alterations felt unnecessary.

Three cards not fitting in the above-mentioned categories were 40, 41 and 43 as can be seen in figure 5.8.

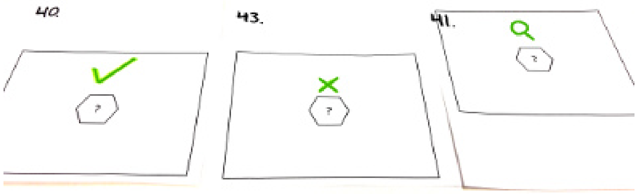

Figure 5.8 Card 40, 43 and 41 are infographics not fitting into the mentioned categories.

The discussions did not reach an unambiguous conclusion of what they meant. Card 40 did not give a clear impression if the operator was to act on the alarm. Several participants were confused by the card and could not figure out what it could mean. Some participants connected the infographic with an approved object. A detected object that did not constitutes any danger.

Card 43 the participants associated with ignoring or delete. If it appeared above an object the object was to be ignored. Similar to what others thought about card 40.

Card 41 was mainly associated with a search- or zoom function. Some participants thought that if they pressed the icon more information would be provided about the object. Others thought that if they pressed the icon a zoomed image would appear allowing them to study the object closer. This, however, fell outside this thesis scope since it was an interactive feature not a *visual presentation of information*.

History

The history cards also naturally got a hierarchy after all test participants voted on their top four. To sort out the most interesting history cards a line was drawn at 4 points. Every card with more than 4 points was regarded, resulting in 4 candidates, see figure 5.9.

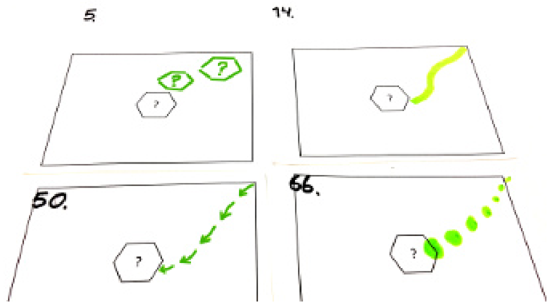

Figure 5.9 Top rated history cards; 5, 14, 50 and 66

These candidates follow a simple pattern, they are showing the past in a simple manner also they do not try to predict a direction. Card number 50 got the least amount of points, concerns were raised about the high cluster of information, every arrow containing information. Card number 5 at third place was visualizing the object in earlier positions. Perceived as a shadowed image of the past.

Card 14 and 66 were ranked as the favorites within the *history* category with 7 points each. 14 showed the past as a continuous line and 66 uses dots in decreasing size from the object. The participants perceived number 66 as a time indication. During discussions, it was mentioned that a similar effect could be accomplished by using colors or fade outs on card 14 and 50 as well to achieve the time aspect. Several participants mentioned speed as an interesting feature when speaking generally. Many participants selecting 66 also selected 14 than vice versa, 66 was more commonly selected as the best possible single solution. Comments on 14 concluded that it was good together with a *speed indicator*, almost identical to 66. Based on the participant's objective data and subjective information number 66 was the concluded history visualization.

One participant thought that the *history* could be unnecessary in the picture, motivating with an example. If the object was seen directly when it entered the area, then the operator would know where it came from and the extra information might only be disturbing. The participant suggested a *toggle function*, where the operator could turn the history trail on or off.

Combining cards

The last step of the exploratory test consisted of combining cards from *Marking* and *History*. *Identity* cards were ignored since they represent specific objects or information. They were bundled together in order to create an easier choice between the two other categories. In figure 5.10 some of the common cards to combine are displayed.

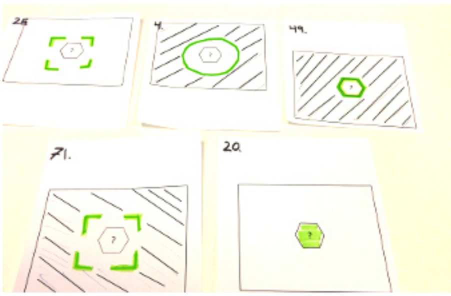

Figure 5.10 Card 25 and 4 or 49 were common to combinations. 71 was an example of 25 with a mask and 20 was a simple marking.

Card 25 and 4 were common in the combination step, they were quite plain, and could be used with most of the *History* and *Identity* cards. Several participants raised concerns about the video being to cluttered when showing the *marking*, *identity* and *history* at once. Therefore the idea to be able to toggle each of the aspects to make a fair comparison in the mid-fi phase were promoted.

Card 25 received 7 points from participants wanting to include card 25 in one or several combinations. Card 25 also received 3 points from the participants wanting it in a combined card. Example, card 71 was a combination of card 25 and card 4.

Card 4 also received a total of 10 points, where 3 points came from card 49 and 3 points came from card 71. Card 49 and 71 are both masked. If the masking was disregarded in the sense that it could be added or removed from any card, the hierarchy would change. Resulting in card 32 being the second best card. Card 32 received 5 points.

Table 5.2 The distribution of points for the top cards, sorted by total points. Example, Card 25 received 7 unique points and additional 3 points from card 71. Card 71 was a combination of 25 and 4.

Card	Points	Information	Total points
25	**7**	Received 3p from 71	10
4	4	Received 3p from 71 and 3p from 49	10
32	2	Received 1p from 51 and 3p from 49	5
66	4	Received 1p from 69	5
14	4	Plain trail	4

The other cards received between 0-3 points each and were disregarded. Card 4 received a lot of points since it is representing a graphical visualization of the masking option and used as the reference to masking the surroundings when combining new cards.

Card 20 was often brought up, but when the participants talked about the object in general, some pointed out that covering the object might be undesirable.

Some participants wanted to use different markings as a way to identify the objects. Marking objects in the video could be done in different ways based upon the identification of the objects. Object information may be visualized at a specific place in the view, mapped to an object. The mapping could be done with color coding or identification markers. Some participants wanted the objects to have unique colors so that the history of each of the object would be easy to separate.

Factors affecting the result

The exploratory testing had a personal approach, resulting in subjective data based on individual thoughts. The biased data was reflected in the structural collection of data. No specific documentation sheet or test questions were planned beforehand. The lack of prior structure reflected the variety of subjective answers. In comparison to a structural question-based approach which could have given a different score. Based on the score, a certain winner could have been found. The structural approach would have been highly dependent on prior knowledge and comprehensive gathering of information and a high number of cards. The subjective approach was more relaxed but required more processing of the result. In this early stage of designing the subjective data gave a basis for creating the complete lo-fi prototype suggestions.

The result might be affected by the test secretaries subjective notes and the test leaders memory. It could have been reduced if the test leader would have written down thoughts after every test and the secretary followed a predefined note structure. Overall the influence from the test leader and test secretary was reduced by the objective data collected in the top four and combination step.

Cards were added along the way by the participants. These cards could not receive points from the earlier participants. Instead, the points were distributed to its subgroups. The distribution resulted in higher points to subgroups even thou the joint card was preferred, and not the sub card receiving the points.

A lined square, *bounding box*, was the current representation in object detection software and many of the participants had seen the bounding box representation before. This might have influenced the outcome of the test. Also, most of the test participants were engineers and the testing might have had a different outcome had non engineers been asked to interpret the feature design cards. The knowledge of implementation and hardware limitations was affecting the outcome, even thou they were asked to disregard these factors. The participants with insight in the industry might relate easier to the end user. This could both have been an advantage and a disadvantage for the test result.

It was planned to include more non engineers as test participants but this was compromised due to several cases of illness. The result could have been more reliable had the test base been larger. However due to time limitations and the size of the thesis as a whole, arranging larger tests and collecting more data was not an option at this stage.

Concluded low-fidelity prototypes

To narrow down the alternatives further the transparent *mask* was decided to be used with all combination later in the mid-fi phase. Other features derived from the initial phase were also used as an overlay of all alternatives later on in the design process.

In this phase, however, the goal was to narrow it down to approximately two prototype combinations. Two marking cards were chosen to represent the test participants top choice, card 25 (Box) and 32 (Contour). The history trail, card 66 were combined with card 25 and card 32 to create the end prototypes **A** and **B**. The figure below 5.11, presents the end result, **A** and **B**.

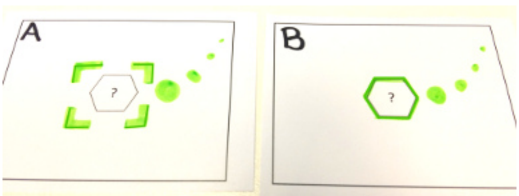

Figure 5.11 The end result of the lo-fi phase, **A** an open box with a trail and **B** a contour with a trail

It was decided that the identity cards, among them the alert triangle, voted most popular would be presented in different ways combined with the four combinations above. This to enable the mid-fi phase to clarify how these should be used later on.

After discussions about enabling the operator to see as much as possible of the camera view it was decided that the geographical visualization, the map, should be moved into a separate window in the interface and thereby fell outside the scope of this thesis.

There were other information that the test participants felt important such as speed and distance, these might be included later on in the process together with the map. These aspects were preferred to be put externally to the video stream in the same manner as the map.

6

Mid-fi prototype phase

In the *mid-fi phase* digital scenarios were produced to enable the test subjects to visualize the design concept. The color, gestalt, and design principles were used as a support in the decision making phase. The information from the *lo-fi phase* together with the *design guidelines* and *design norms* created the base for the *hi-fi phase*. During the mid-fi phase, two tests were performed to help decide the conclusions.

6.1 Graphical visualization of the low-fidelity result

Axis has a set of rules and design guidelines when it comes to interfaces and design. The prototype design took them into account in the design process. Most of the guidelines and rules were applicable to interactive interfaces such as styles for buttons etc. However, some things affected this design process. There were predefined colors that were to be used in the design to make the video stream and interface coherent.

Another design choice that was made were to incorporate *material design*. It was used in accordance with the rules and design guidelines, examples are presented in figure 6.1 below. Material design is a project developed and maintained by Google. It is presented as a visual language created with the intent to use classic principles of visual design and add a sense of modernity and innovation, [Google, 2017]. Standardized icons are an example of a benefit by following the material design in the lo-fi prototype phase.

Figure 6.1 Example of Google material design symbols

The colors chosen to be used throughout the design were *red* and *yellow*. They were already associated with different severeness of threat levels which corresponded to the information wanted to be conveyed with the colors. A triggered object was marked yellow and then if the software manages to classify the object as a real threat the alert symbol would appear and the yellow would turn into red.

Marking

The decision to use *identification* and *marking* together was strengthened by the *color principle*, saying that the marking should not be done solely by coloring. Therefore all combinations were tested to see which combination was preferred.

Another principle said that if the marking was too thin it would be hard for the visual perception to distinguish it from the video stream. Combined with the aspect that the end user wants to see as much of the video stream as possible, a rather thin line was preferred.

A combination of dark and light colors was decided to be used since it improved the contrast in a noisy environment.

Both Norman's *design principle of consistency* and the *gestalt principle of similarity* were applied when objects were marked the same way. Forming a group consisting of all objects triggering alarms. This did not necessarily include the deciding on the coloring or sizing of the markings. The size could be altered to imply more important triggers such as humans or human activity.

Coloring the Marking The coloring between triggered objects had to be considered too. If it were decided to give all objects a distinct color, then the video stream with several objects could be perceived as cluttered. On the other hand, if all markings had the same colors it could be hard to distinguish which history trail belonged to which triggered object if there were multiple objects moving closely together on the screen. Assuming the history was presented in the video stream. The marking was decided to have one of two colors, red or yellow.

The marking consisted of two parts; the outer shape and the inner shape. The outer shape had a compact color, red or yellow, the inner shape was black as can be seen in figure 6.2. This so that the markings would be easily distinguished regardless of the background.

Figure 6.2 Mid-fi visualization examples. Left: A box marking with masking. Right: A contour marking.

Contour This marking consisted of the same two parts as presented above. The main difference was that the lines for this alternative followed the *contours* of the detected object. Both the color and the fact that the object and its marking were likely to move in the video stream helped the user to focus on the object.

Box This marking related closely to the *gestalt principle of closure*. People tend to see the open square as a *box*. The box did not cover as much of the background compared to a *bounding box*. The size of the box was decided in accordance with the *proximity principle*. If the box was big and the corners short the corners would be widely separated and they would stop being seen as part of a square. Instead, it was decided that the corners would be approximately one-fourth of the shortest side of the box. Also, a minimum and maximum size for the box would be set to keep the right proportions.

Identification

Based on the testing up till now there were three separate ways of showing the identifications. Since the identifications were not the prime focus of the lo-fi phase, later tests would have to reveal which way was preferred. Still, the knowledge received from the lo-fi phase was that *infographics* were preferred over text, which was consistent with, and strengthened by the theory. Examples of the lo-fi identifications symbols are shown below in figure 6.3. The identification symbols had to be designed in a way that did not invite the end user to interact with them. In a way, this reflected on Norman's *principle of affordance*. The outcome of clicking a symbol shall be nothing, and the symbol shall hint it, not invite to interaction. In a way, Norman's *constraint principle* also applies here since the end user should not be able to make a mistake using the visualization. This since there is no interaction with the visualization and the end user is not invited to interact with the visualization.

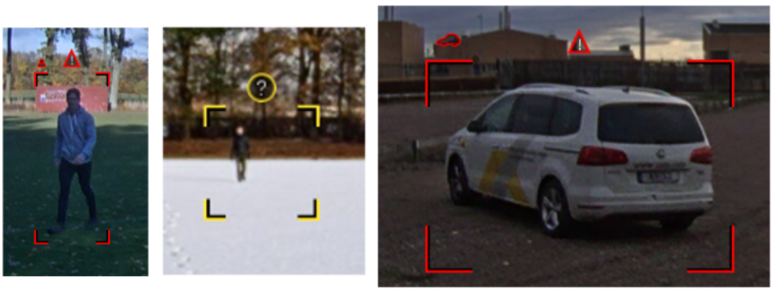

Figure 6.3 Left: *Alert* and *Human*. Middle: *Unknown*. Right: *Car*, an alternative to the human identification symbol.

The symbols were decided to be the same for all tests even though the symbols were placed at different positions in the video stream.

Alert The preferred *alert infographic* had the triangular shape of a warning sign with an exclamation mark inside. When designing the infographic it was decided to follow the gestalt principle of similarity. The alert symbol had the same color as for example a *human*, or a *car*, symbol as well as the same color as the marking as they should be thought of as belonging to the same group, which was the desired effect.

The alert symbol consisted of three separate parts, the triangular shape, the red color and the exclamation sign. All three conveying the same message of warning.

Unknown When it came to the question symbol, an *unknown object*, most of the arguments for the alert symbol could be applied. This question symbol was wrapped in a circular shape to prevent it from blending with the background. This also highlighted the fact that the system did not manage to identify the object and that the operator should look closer at it. It was not as critical to chose a distinct color for the question symbol since it did not state a warning rather that tell the end user the object was unidentified.

Now three separate identifiers were used together; the circle, the question mark, and the color. The question symbol was decided to be a yellow outer line in the shape of a circle. The circle would be filled with a transparent black color and the question mark would be concrete white. This enabled the symbol to be easy to distinguish from a, light as well as a dark, background.

Human The human symbol was chosen to be an infographic of a human instead of a textual alternative. The symbol was thought to help the end user clarify that the system had identified a human or in the future a car or similar. It was also important to remember that the system never could be 100% sure, perhaps an indication of confidence might be needed. This would have to be outside the video stream though along with other information that could be useful for the end user.

A human and a car symbol were created to provide the end user with correct information for the tests. It was decided that the human and car symbols would be filled with transparent black inside yellow or red lines. The color of the lines depended on if the object was identified as a threat or not.

Identification positioning There were several aspects of where to place the identification symbols and the different alternatives were the following.

Identifications one (*Id 1*) The identification symbols for Human, Alert and Unknown were placed right above the targeted object. The colors did not have to match the marking of the object for it to be associated with the targeted object. This due to the principle of proximity.

Identification two (*Id 2*) The identification symbols for Alert and Unknown were placed in the top left corner of the screen. The Alert symbol was colored red and the Unknown symbol colored yellow to indicate the severeness of the symbols.

The symbols were combined with the color of the markings. If the software detected a human activity then the alert symbol appeared in the corner and the marking around the specified object turned red. If the software took the decision of being unsure of a triggered object the yellow unknown symbol would appeared in the top left corner and the object marking stayed yellow. To separate the symbols and placing them all in the same place would give the operator a possibility to quickly get an overview of which type of objects were detected within the video stream.

Identification three (*Id 3*) The identification symbols for Alert and Unknown were to be used and placed right above the targeted object. Here the principle of proximity gave the end user a sense of the identification belonging with the targeted object. Therefore the coloring does not have to be the same.

History

The *history trail* consisted of dots in different size trailing behind the targeted object indicating where it had been. There were two choices when it came to color. If the history trail had the same color as the marking of the triggered object it would make it easier to distinguish each separate object according to the similarity gestalt principle. Another choice was to use the coloring as a speed indicator. For example, if the object was moving fast then that trail part could have been red and when it moved slower the trail could have been green.

In an early design iteration step of the history trail three issues were found. The first issue was that a trail would clutter the view, the dots would be preventing, shadowing other objects. The second issue was due to the camera angle. The camera view would most likely be seen from above hence making it hard to create a viable solution for displaying the history trail in a natural manner. The third issue was more of a solution, it consisted of creating the history trail in a *map view*. Based on these issues and solutions it was decided to leave history out of the mid-fi testing and include it in the map view instead and add a toggle function to it.

Masking

Adding a transparent *mask* over what should be perceived as background followed the *gestalt principle of figure/ground*. The smaller unmasked area would automatically be perceived as the figure and thereby the area of importance while the masked area would appear as the ground, background. The masking would also help to provide a better *UFOV* since a considerably smaller area needed attention and the cognitive workload would decrease. An example of the masking functionality can be seen in figure 6.2.

Toggle function

Even though the *toggle function* was an interactive function it contributed to the design concept as a whole. The transparent masking and history trail could be toggle alternatives since it was not sure the features were wanted as a standard option. Masking could potentially prevent the operator from detecting objects in the video stream. By darkening the image the operator had a harder job to detect objects that the software potentially missed. Positive aspects such as helping the operator to focus on the critical area could also be seen. Therefore having the masking as a toggle option in the mid-fi would help to find out if the test participants preferred the usage of masking or not, when they could see what it could look like on a simulated background.

6.2 Mid-fidelity prototype creation

The main idea for creating the mid-fi prototypes was to use graphical editors to add features on different layers and in that way be able to combine the features in different ways and save the alternatives as separate images. Scenarios where used to present the features in different environmental settings.

Environmental scenarios

The mid-fi prototypes were based on 3 plausible scenarios, created based on information received from camera usage experts during the lo-fi phase. By stating specific scenarios, the different visualizations could be tested comparing weaknesses and strengths. The scenarios were based on 3 still frames and tested edge cases that could potentially cause problems for the design solutions.

Scenario 1 The first scenario showed two people walking on each side of a tree. The goal with this scenario was to show the differences between when someone was detected within a *trigger zone* area versus outside a trigger zone. In the scenario, it was assumed that the right most person was inside the trigger zone and the left most person was outside, where the movement was allowed.

Scenario 2 The second scenario showed a person and a dog walking in the dark where it was harder for the operator to distinguish what was happening in the scenario. This scenario aimed to show how it could look when detecting two different kinds of objects in two places of the screen.

Scenario 3 The third scenario showed two persons, one hiding behind the other. This scenario aimed to show individual visualization when the persons were grouped.

Combined features

Based on the features presented above these were to be combined into images showing all possible combinations of markings and identity together with and without masking. The combinations can be seen in table 6.1. Some examples can be seen in appendix A.

Table 6.1 A list of the 12 combinations and the features they consist of.

Combo	***Masking***	**Marking**		**Identity**		
		Box	*Contour*	*Id 1*	*Id 2*	*Id 3*
1	*x*	*x*		*x*		
2	*x*	*x*			*x*	
3	*x*	*x*				*x*
4	*x*		*x*	*x*		
5	*x*		*x*		*x*	
6	*x*		*x*			*x*
7		*x*		*x*		
8		*x*			*x*	
9		*x*				*x*
10			*x*	*x*		
11			*x*		*x*	
12			*x*			*x*

For each scenario, there would be 12 different combinations to cover all possibilities. Multiplying the combinations with the three scenarios resulted in 36 images. Images were created to present all combinations within all scenarios. The different design concepts were represented and tested, none of the scenarios could be removed without loosing an aspect of the testing.

6.3 Mid-fidelity prototype test

To avoid biased participants a decision was made to use test participants not working with object detection or graphical visualization. The first test was performed in person with over 45 students at the Faculty of Engineering, Lund University. The second test was spread through social media to collect a broader collection of data and received over 180 answers. The goal of the tests were to inquire which combination was favored, based on an operator situation, fast identification and accurate identification. The combined result became the base for the hi-fi prototype.

Mid-fidelity test 1

A simple java application was created to assist in performing the tests. The application displayed two randomized combinations in parallel, as can be seen in figure

6.4. The three scenarios were displayed in three rows, one row per scenario. The test participants compared the combinations and were asked to favor one column, left or right. The comparing phase was repeated enough times for the participant to evaluate each combination twice.

Figure 6.4 The test program comparing combination 10 and combination 4. The test participant can choose what combination is preferred with the buttons *left* or *right*.

In figure 6.4, the difference between the combinations were the masking. Based on the images the test participant had to generate an overall opinion of which combination worked best for all three scenarios. Each test consisted of 2x6 comparisons with randomized combinations from combination 1 to combination 12 and was estimated to take 5 to 10 minutes.

Result of the first mid-fidelity test

The testing indicated that the Masking aspect was both visually aiding and distracting. Masking could be a help for AO when an alarm suddenly appears to highlight the area of interest. However, for a SO who is watching the video stream continuously, it could be distracting. Masking was mentioned as preferred in the snow scenario, but not in the dark scenario. Many of the test participants found both issues and benefits with the masking. A solution could be to use an opacity adjustment slider to enable the operator to adjust the mask or remove it completely. The masking is henceforth outside of the scope of this thesis but might be incorporated in the hi-fi prototype to show the concept.

Overall the result was ambiguous because of the test setup. It was hard for the test participants to establish an opinion that stayed consistent throughout the test. As an example the first 3-4 pictures could display the same marking option, not giving the test participant the opportunity to disregard that marking option completely. The problematic with this was that it was all randomized and several aspects could be compared at the same time. For example, when both masking and marking was compared the problem occurred when the participant wanted to choose left because of the marking aspect but also wanted to choose right due to the masking aspect.

It was, however, possible to single out the comparisons where only one aspect was tested at once and these comparisons gave a valid indication. These comparisons were noticeably fewer than the total amount of comparisons performed.

As for the comparison between combinations where the only difference was box and contour, 10 people preferred box versus 11 for contour. Counting only these comparisons the answers were too few to be the sole base for a decision of which combination should be used as a base for the final mid-fi prototype.

The evenness was repeated when the identifications were compared, as an example Id1 compared to Id2, where Id1 received 7 votes versus 8 votes for Id2.

The result of the first test did not create a clear combination winner. Some aspects that might influence the result could be diffuse instructions, unnoticed changes and transfer of learning effects. Perhaps the test should have been complemented with an interview to gather the thoughts during the test process. Another way to go was to create a longer test or force the test participants to rank a favorite by displaying their result and let them compare their opinions. A second comparison test was designed to provide further data to base a decision on.

Mid-fidelity test 2

In the second test the alternatives of not using marking or identification was added to give a broader view of if the marking and identification were contributing to the user experience or not. Leaving the test participants with 24 different combinations to choose from.

The second mid-fi test was designed to encourage the test participant to design their own preferred experience. Based on a strict use case, explaining their role and situation.

An online form was created and distributed online, displaying the different choices individually. Then allowing test participants to design their own preferred experience by choosing; one of the marking alternatives, one of the identification alternatives and one of the masking alternatives, see figure 6.5. After this, they could visualize their choices with the different scenarios, see Appendix B.They were then allowed to submit their choices or go back and redesign. The process permitted the test participants to be creative and explore different alternatives in their own pace.

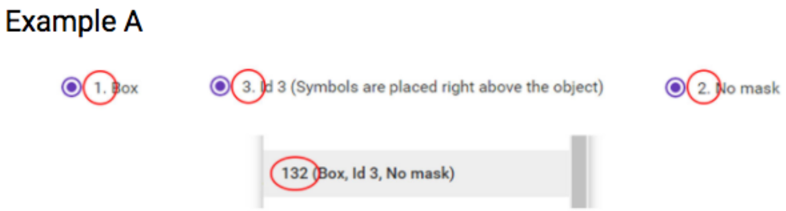

Figure 6.5 Example of how to combine your preferred type of marking, identification alternative and masking option to answer the online survey.

The second test was distributed through email and shared on social media, the simple design of the test and the ability to share online allowed more than 180 participants to submit their preferred combination.

Result of the second mid-fidelity test

Concerning the marking alternatives the box was the preferred marking option. The box received 56.4% of the votes compared to 43.1% for the contour. The missing percentage went to the alternative of not using a marking. The box was the most preferred marking option when comparing identification options internally as well. *Id 1* was preferred over Id 2 and over Id 3. The percentage of the total votes for the different identification alternatives can be seen below in table 6.2.

Table 6.2 The percentage distribution for the **identity** alternatives received from the second mid-fi test.

Identity			
Id 1	*Id 2*	*Id 3*	*No Id*
63.5%	13.8%	19.9%	2.8%

The overall combinations which received more than 12 votes are presented in figure 6.6.

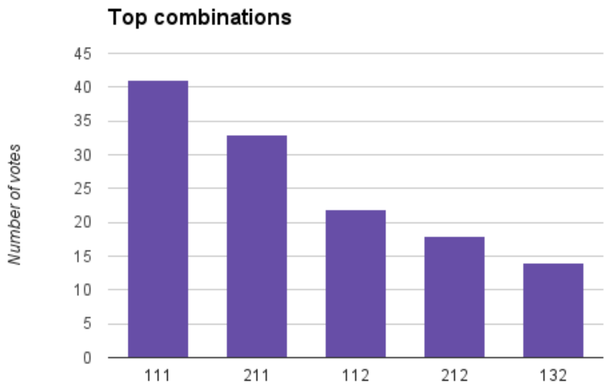

Figure 6.6 The top voted combination from the second mid-fi test.

The result was that a total of 22.6% favored the specific combination 111 (*Box with Id 1 and a mask*). And 18.2% favored the specific combination 211 (*Contour with Id 1 and mask*). These two were quite even, but considering the result when comparing all box combinations versus all contour combination, the box was preferred. Altogether the result of the second test was, combination 111 (*Box with Id 1 and a mask*).

Figure 6.7 Left: 211 (*Contour with Id 1 and mask*). Right: 111 (*Box with Id 1 and mask*).

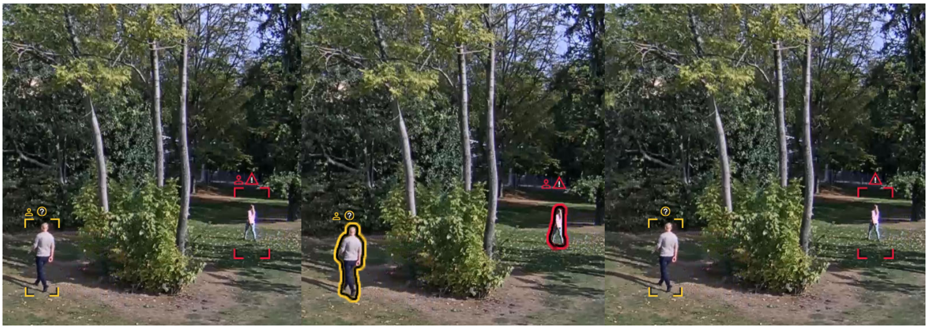

Figure 6.8 Left: 112 (*Box with Id 1 and mask*). Middle: 212 (*Contour, Id 1, No mask*). Right: 132 (*Box, Id 3, No mask*).

Conclusion of the mid-fidelity test results

Several things were established during the mid-fi phase. From the first test the masking aspect was decided to be an interactive function, using a slider to determine the opacity. This way the masking was excluded from this thesis, still a mask will be used to showcase the intent of the masking function.

After the second test, marking with an object with a box was preferred over using a contour. Id 1 was preferred over Id 2 and Id 3. This meant that the concluded combination resulting from the mid fi phase was combination 111 (*Box with Id 1 and mask*), see figure 6.9

Figure 6.9 Combination 111 (*Box with Id 1 and a mask*), the final mid fi prototype.

The combination was used as a base when developing the hi-fi prototype in the next chapter.

7

Hi-fi prototype phase

In the hi-fi phase, a *digital prototype* incorporating the visual design concept was implemented. The hi-fi prototype was based on the final still frame mid-fi prototype, *combination 111*. The hi-fi prototype was lastly *validated* by end users.

7.1 High-fidelity prototype creation

To give the hi-fi prototype a genuine appearance the *company's design guidelines* was applied. They were central in the process when it came to e.g. determining *colors*. The hi-fi prototype was coded in C with additional support interfaces. The hi-fi prototype program was a program that received input data in form of a video stream and a stack of information files in *XML* format, the information was interpreted, synchronized, combined and visualized.

Input data

The input data consisted of two parts, a stack of XML files and a *stream of images*. Synchronized and combined, an XML and an image, created a new frame including both the original image and the design drawn on top based on the information from the XML file. The XML files held the information about what was present in a specific frame. The XML information was generated by extensive algorithms and decision making, which was not a part of this master thesis.

Some of the XML files had to be manually adjusted since the software was not developed to be able to make the decisions that were requested by the workshop result from the initial phase. The design was dependent on the software being able to separate humans from other objects which at this point of time was not doable. The alterations of the XML files were made to include such specific information as fabricated classifications. The XML files and the images were synced, combined and then presented.

Tools and techniques

As mentioned, the hi-fi prototype was a software program written in C. The program used an XML parser to interpret the XML's. To display and alter the images several Application Programming Interfaces (API's) were used. An API called Gimp ToolKit (*GTK+*) were used to display the image window on the screen. An API called *Cairo Graphics*, a library used for 2D vector graphics were used to draw on the original image.

To decompose the XML files a free to use XML parser was used, [Veillard", 2017]. The parser was altered and modified to gather the specific XML information needed for the design. As an example the XML included information about the classification of an object, this information supplied the program with the information needed. With the parser had extracted the information it could be forwarded into Cairo to initiate the design process. Cairo drew the design which was then presented as an overlay on the original image. Displaying the combined image on the computer was enabled through GTK+. The original image was stored without any excessive information, as it was important that an unaltered image could later be used in court.

To create a genuine sense of real time experience, the program displayed the output with a speed of 5 frames per second (5 fps), displaying a video together with the design drawn in an overlay.

Alterations applied during the high-fidelity phase

During the implementation process, there were several different types of XML files, described below.

- An empty XML file, no movement in the scenario.
- Only radar, motion detection information, radar data. No visual confirmation of movement, for example, a dark night.
- Only MOTE, motion detection information, alterations in the picture. Visual confirmation of movement or false triggers like shadows moving.
- Combined information from radar and MOTE.

In the case with an empty XML file, no alterations were made to the frame overlay. In the case where only radar was accessible, it meant that the video had not detected the movement. Since it was not possible to pinpoint a *radar distance position* in the *image*. The classification could not be done without visual confirmation. The design of the only radar case was to keep the design consistent with the other cases. As familiar as possible to create a sense of consistency within the program. The visualization was the same except not showing the identification and classification symbols and the height was the whole image since there was no information to base a decision of classification nor height. The case with only MOTE was regarded as a normal case, without distance information.

Combined information was the most common case, hence it was also the case with the highest design focus. As an example, the *thickness* of a marking line was changing depending on the *distance*. When an object was far away the lines were thin, this enables the operator to see more of the surroundings. On the other hand, when an object was close to the camera, it would be easier to locate and hence did not need thicker lines. To create a *dynamic visualization* the thickness of the lines changed, so that an object that was close had a thicker line. But the changes were small, it gave the effect without cluttering the image. The information could be gathered from the radar and when visualized it gave a sense of three dimensional for the objects.

To keep the clutter of the image to a minimum it was decided that when 2 objects encountered each other's markings, only the marking of the object closest to the camera would be displayed. Fewer markings in a small area gave the operator an easier task in interpreting the video. Also, the operator's focus would already be in the right place.

The size of the mask was decided to be larger than the marking providing the operator with an overview of what happened in proximity to the triggered object. Enlarging the uncovered area gave a sense of the identification and classification symbols being a part of the triggered object as opposed to the mid-fi prototype where the identification and classification symbols were drawn on top of the dark mask. This sense was strengthened by the *gestalt principle* of *proximity and similarity*.

7.2 High-fidelity prototype validation

To validate the hi-fi prototype two validation tests were performed. One with a *technical development of video and remote services manager*, from here on referred to as *manager*. And one validation test with an *AO*.

A *test plan* was created to structure the testing and ensure the extensiveness of the *hi-fi prototype*. The purpose was to *validate the features and visualizations* that created the prototype design and evaluate if it were satisfactory and fulfilled the needs of the end user. After the demonstration, the test participants were asked a couple of questions to conclude the testing and ensure that the interviewees were satisfied. The validation test also provided an insight into which parts of the design that could be useful and which parts would most likely not be an improvement to the existing system used. The hi-fi validation plan can be seen in the appendix C.

Process of demonstrating the high-fidelity prototype

The hi-fi prototype was demonstrated at an end user's office. The output from the hi-fi implementation, images with the design overlay, was recorded into 6 videos presented to the end user. The demonstration included watching the 6 videos that visualize the different situations and features from the prototype design. Before demonstrating the videos a couple of questions were asked, they were open questions and encouraged the interviewees to elaborate and explain how their working situations are as of today. The 6 videos shown were:

Video 1 Two identified persons were marked as humans, crossing each other's paths.

Video 2 An identified car was driving in a parking lot.

Video 3 A single human was walking towards the camera. At first unknown but later identified.

Video 4 A night shot, a person was crossing a golf range. The radar located the person, later the person was lightened up by a flashlight.

Video 5 A night shot, a person was crossing a golf range. The radar located the person, later the person was lighted up by a flashlight. Also, a mask was added.

Video 6 An identified car was driving in a parking lot. Also, a mask was added with an opacity slider function.

High-fidelity prototype demonstration

The observations from the two hi-fi prototype demonstrations are presented below. The result was discussed and evaluated to validate the design process and the thesis result and purpose. The demonstrations were held separately.

The first high-fidelity prototype demonstration Demonstration with the manager. The participant was not directly working with the *surveillance systems*, but rather with the end user. The participant often mentioned that he represented his own personal thoughts and stressed that an operator may answer differently and that their opinion should be valued over the participants own opinions and thoughts.

Before the demonstration, the participant explained that what they used today was a software that put a *bounding box* as marking around the triggered objects. A *bounding box* is the smallest box imaginable based off what the system consider the size of the triggered object. It often draws the box partially on top of the triggered object and the box in question consisted of a thin line. The participant said that a *bounding box* or similar is extremely important to be able to save time in finding the object for the end user. The time saved by finding the object fast was without a doubt higher prioritized than visualizing everything (no box). Perhaps there could be a *box toggle*. A bounding box or something similar was presented as a basic requirement, a *nonnegotiable demand*. Another demand was the limitation of the decision making, the operator should not have to take uncertain decisions.

The participant was asked to abbreviate on functions and features that could be appreciated by the end user. The participant mentioned classification but was unsure if the operator would need it and said that it depended a lot on the preciseness of the software. An operator would rather have *no classifications* than a wrong classification. *History trails* and *trigger zone* was also mentioned as things that could be of visual interest for the end user. The participant often returned to a specific question, "Does this add value for the end user?".

First, the participant watched video 1 and 2 showing how the *alert symbol* and the classifications, *human/car*, could be used. The first thing that was said after Video 1 and 2 was that the classification might be useful in some circumstances and scenarios but all depending on the *confidence and reliance* on the software system making the decision. The *alarm symbol* was also thought of as redundant since all alarms were treated like threats, no matter what triggered. The participant said that, when an alarm arrives, it could take some time before someone watches the video stream. Then the object that triggered might have left the screen.

After watching video 1 and 2, the participant was asked to watch video 3 where an *Unknown* person walks towards the camera. The response to the Unknown identification was that when an alarm arrives at then it was already considered as a threat. With that said the Unknown identification symbol and the *yellow* color would be redundant.

Then the participant watched video 4 and 5, displaying a very dark shot and a *radar object detection*. The participant disregarded the solution with the motivation that the end user could not make a decision if there could be no visual confirmation of the triggering object. The most important thing was that the end user could react to a *positive alarm*, where a response was needed, and *disregard false alarms*, where no action was needed. Therefore all decisions had to be based on a *visual confirmation*.

Video 6 showed how a mask could be used with an *opacity slider* to help locate the triggered object. The participant was unsure about the opacity slider and masking as a whole but yet again that the real end user might think differently about it.

After every video was shown, a brief discussion took place and the usage of *maps* was mentioned. The participant said that maps would be a huge improvement for the end user. To be able to localize the camera and put the scenario into a context would be very helpful when, for example, directing a *security guard* arriving at the triggered alarm location. As well as locating other cameras or suspects.

The second high-fidelity prototype demonstration The second demonstration participant worked as a *AO*, alarm operator. The participant reacted to alarms and followed a framework to handle the different types of alarm that occurred.

The participant described that on a regular basis they received alarms constantly. They had to react to them fast and accurate, sometimes only seconds could be spared to determine if an alarm was false or positive. Over 90% of the alarms were false alarms, they could be caused by wind brushing the trees, causing movement, or animals running within the trigger area. The participant mentioned that today a *bounding box*, a small red box that draws attention, were used on some alarms but not all. The *bounding box* helped the participant to focus on the right place, but after you had located the triggering object then the box may be superfluous. The ability to remove the box was requested, and further strengthened by a crowded alarm scenario where too many boxes would clutter the view.

After watching video 1 and 2, videos using the *red* boxes and the alert symbol accompanied by the *human/ car* symbol, the spontaneous thoughts were that the box was too big, the function behind the exclamation mark was ambiguous and the Human/Car symbol was unnecessary. With that much information, a frame would be too cramped and obscure. The end user was trained to differentiate between objects on an ambiguous image. Even if the system could detect and classify objects then it would probably not be adding anything since the operators could do it fast enough.

With today's system, it was hard to see what happened if the object was far away, not only since it is graphically small in the frame but also, the *bounding box* is drawn too close to the object which in some cases hides the object behind a non removable red square.

The participant figured that it might be better with a *dot* centered on the object. The participant acknowledged problems that would arise when identifying behavior or action. The dot would always cover the object and therefore the marking option would need a *toggle* function. A situation referred to as similar was, a thermal camera video. The shot is dark, and the object triggering is lit up, but still unidentifiable.

The participant noticed benefits with the *open box* versus the closed one as in the design presented. The open box would allow them to see more of the screen than the *bounding box*. The open box also had space between the object and the markings compared to the bounding box which does not. Hence the participants first impression that the box was bigger than the bounding box. The open box allowed the operator to perceive more information. It made it easier to see what was happening, especially in arms height, when the triggered object was picking up things or doing something.

The participant was asked about whether or not it was a sufficient way of showing distance by adjusting the thickness of the open box. The participant mentioned that a number would probably be more useful since it would give a precision in meters instead of a brief sense of distance. An example of when this could be useful was when the operator had to guide a guard on the ground and it the operator could say *"The target is hiding 20 meters north-west of the container"* instead of saying *"The target is hiding left of the container, in the middle of the screen"*. In these situations, it would be beneficial to be able to give better directions strengthened by the new information.

The operator said that this option would have to be able to be turned on and off if needed since it would be too much information if it was constantly visualized.

Video 3 presented a man walking towards the camera, the box was *yellow* and an *Unknown* symbol was displayed, when he came closer the box turned red and an *Alert* symbol was displayed together with a *human* icon. The participant interpreted the situation as if at first the man was close enough to trigger, but outside of the *trigger zone*, and that was why he was yellow. When he then entered the trigger zone, the zone where he was not allowed to be in, the marking turned red. The *Unknown/Alert* symbols were disregarded since the color indication was enough. Today the operator can not view the trigger zones in the video, they must use additional software. Using colors would improve the system since they would directly receive the information. If they want to see the trigger zone of a specific camera they had to open another program and then search for that specific camera and then they could see how the installer had set up that specific trigger zone. Witch adds a heavy *cognitive workload* on the operator.

Video 4 was dark, an example video that was not representing a real situation. There were never any sites that were pitch black. Either they were lit up or thermal cameras were used. The verdict was evident, if the operator can not see anything then they can not make a decision as if to send a guard or not. Video 5 and 6 presented the usage of the *mask*, the first thing mentioned was that the participant still wanted to preserve the original image, the mask function darkened too much unnecessary. It was not something the participant considered beneficial.

The *map* overview with *history trails* was presented for the participant. It was regarded as a good feature, much welcomed by the operators. Mentioned strengths were the easy guidance of ground personnel and quick geographical overview. Today the guidance of personnel was one of the hardest tasks, to communication correct information fast and accurate.

Another feature to add to the map was the possibility to visualize the trigger zones on the map. The map was considered good as long as it did not cover the actual video. Overall the participant wanted to minimize the number of clicks and interactions with the system, to be able to make accurate decisions fast were prioritized since a large number of alarms were always a problem.

Conclusion of the high-fidelity prototype demonstrations

Some of the features were disregarded and some where welcomed by the participants. An open box was preferred over a bounding box. A map feature was welcomed and sought after. The usage of different colors, yellow and red, were approved by the operator only. But the operator associated the yellow color with an object outside an trigger zone. Another interpretation was that the object was unclassified by the system.

After the demonstrations, the problem of the operators not being able to see the trigger zones were obvious and the concept of colors reflecting that aspect was highly preferred. The manager thought that the operator would not be interested in anything that was not already red. But the operator mentioned that the yellow coloring would be a helpful feature. It shows that even within the same company the thoughts were widely spread about what feature was useful.

In this situation, the opinion of the *operator*, who is the defined end user for this thesis, weighs higher. This was also supported by the first demonstration where the *manager* said that he could not know for sure what the operators would like. The answers were professional guesses.

Among the things disregarded by the end user was the *Unknown/Alert* symbol. Since they did not add any additional information if the system was color coding in *yellow* and *red* already. The *classifications* were disregarded since the end user often viewed from afar, in that case, the symbols would be considered to clutter the image. Near the camera, classifications would lose their purpose since there was no difficulty in classifying objects near the camera. Both the operator and the manager agreed that the symbols were unambiguous. By removing these symbols the *cognitive workload* for the operator would decrease, strengthen by the theory, the *UFOV* and the detection of objects in the periphery would increase. The result is an easier cognitive workload, reducing the tenaciousness of the operator.

Removing the *Unknown/Alert* symbol disregards one of the general principles of *graphical visualization* that says that, colors should not be used alone to convey information. Colors should be used together with positional information, labels or other cues to enable someone who could not distinguish color to interpret information.

When regarding the *masking*, they were both slightly dissenting, it would not help the operators in their work today, hence the mask was disregarded.

A thing they both agreed on was that the possibility to detect an object with only radar was not an improvement for the operators since they needed *visual confirmation* before acting. An alarm where the camera video was dark and the radar detects movement was not enough for the operator to act on. Therefore, the only radar scenario was disregarded. A radar could be used to detect movement and then switch a lamp on.

They both agreed that a *map* with *history trail* would be appreciated, the operator added that if the *trigger zone* also were marked on the map if would be even better. It also became clear that a feature showing the *distance* towards an object would be appreciated and when using numbers it is important to remember the general principle from the theory chapter to use whole numbers e.g describing an object as *20* meters and not *20.00* meters away from the camera. Also that the labels should be presented horizontally and not vertical to facilitate the end user's workload so that UFOV remains as large as possible.

Verifying the validity of the result The surveillance operator was a single person, with personal ideas and thoughts. The subjectiveness of this person was weighted versus the features and design choices that others, who has been contacted through the process, came up with. The AO was presented with the result of the process and viewed the result with eagerness and with a critical standpoint.

The AO was not against change, but was critical of changes, since the result would potentially become a part of the day to day work. The aspects were viewed from different viewpoints and factually based opinions were voiced.

The AO had never been presented with the concept of *combining radar technology with camera surveillance* and was not fully introduced to what new data the radar could provide. Probably affecting the AO to not being able to give more than spontaneous answers to the questions involving the radar cases. AO's critical attitude is highly important in concluding the hi-fi prototype. The attitude enables the accuracy and authenticity of the end result.

High-fidelity prototype conclusion

The concluded hi-fi prototype differs in several ways from the mid-fi prototype. The changes involved disregarding several aspects, the disregarded aspects are presented below:

- The *Unknown/Alert* symbols.
- The *Human/Car* classification symbols.
- The distance indication by *box* thickness.
- The *mask* feature.
- The case with only radar information.

The concluded hi-fi prototype, is presented in figure 7.1.

Figure 7.1 Consider the right side of the bush is the trigger zone this is what the design would look like.

The design consist of the following:

- An open *box* marking.
- Colors, *yellow/red*, for showing if the object is outside or inside a trigger zone.
- A *map* that shows trigger zones and the *history* of objects.
- An option to toggle an objects distance information.

8

Discussion

This chapter discusses the whole process divided into sections *Defining the end user*, *User-centered design*, *Iterative design process* and aspects that could need future studies are presented in the *Future work* section.

8.1 Defining the end user

In this thesis there were no specifications defining the *end user* given at the start, therefore the result is based on knowledge gathered through scrutinizing the *sales model*, investigating the *work environment* and *proxy interviews*. The proxy interviewed employees were not working directly with the end user, instead, they were working with the end customer. Making them the ones most knowledgeable about the end user. The proxy information step could be the reason why the end users found some *features undesirable*. If precise data regarding the user had existed then the result would most likely be more accurate. Precise data such as personas or user journeys would have been useful, but were non existent. The lack of access to the end user and the lack of time made it difficult to create personas and user journeys. One thing that was discovered was the lack of knowledge about how the end customers actually work. Few of the proxy interviewed persons had seen the end user at work, in their work environment.

Complexity of the defined end users

Since this thesis focused on the common ground features that would work for both AO and SO. It could have been easier to target one of *AO* or *SO* as the end user. If end user tasks has varies greatly, targeting the real end user would demand an excessive amount of data and analyzing but could favor *future design processes*. In short, being able to see what differs between the tasks of an AO and a SO would have contributed a lot to the discussion. As of now only one side of the AO and SO was heard and seen and the results were all based on the AO point of view. After visiting the work place of the AO it was clear that to some extent they were watching the videos live. In the manner the concluded SO would do. Showing that the difference between AO and SO was smaller than realized. This could mean that they would have a lot in common and the fact that only a AO were *validating* the design might not alter the results.

Several aspects that fell *outside the scope* for this thesis were found during the process. Features that could benefit AO or SO or both in the future. Such features were both a part of our result but also subject for future study and therefore these are all described under *future work*.

8.2 User-centered design

A user-centered design process was performed even though the actual end user was not involved in the process until the validation phase. Instead, design decisions were based on the result of several test results and *design principle* throughout the design process. Based on the presented information the test participants created their own *cognitive picture* of the operator. These cognitive pictures varied depending on the participant's different backgrounds and experiences. In the long run, this means that there is a chance the participants creates a correct cognitive picture close to the real end user. However, there is also a risk that the participants does not think like the end user, leading the design process astray. In the worst scenario make the design unsuited for the real end user.

As an example, it could be hard for the mid-fi test participants to imagine a work environment and working with the software daily. In the test situation, when they only have five minutes to decide which combination they prefer, a misconception might occur. A real end user might be able to see features as excessive while for a test participant the same features could be essential. Therefore it would be preferred if the end user were involved in the initial and final phases of the process. Lacking this possibility, a great number of test participants were testing the mid-fi prototypes to achieve a more representative result. The idea was to create a big variety to be able to generate a clear pattern of which combination was preferred.

8.3 Iterative design process

Working with a design process including three *prototype phases* helped the design process in several ways. Since there were a great number of features that could be of interest. A low-fi prototype phase to find the most important features was necessary. The fast and repetitive phase enabled features to be discovered and evaluated early on, theoretically reducing the time spent on features that could be disregarded in later phases. In our case the lack of precision when defining the end user lead to time spent on undesired features.

Using the mid-fi phase with alternated images gave the test participants a fair chance to evaluate which visualization was preferred for the different features. Without the hi-fi prototype having to be implemented for all 24 alternatives represented in the mid-fi phase. The mid-fi phase contributed with making the big design decisions, for example, choosing between an open box or a contour, in an easily changeable environment.

There were a lot of aspects affecting the design that were merely visible when a hi-fi prototype video was studied. The hi-fi phase contributed with the detailed design decisions such as thickness of lines, fonts and specific colors. Being able to separate the big decisions from the detailed decision enables focus to be put on all aspects but at different stages of the process. With the time limitations for this thesis using more than three prototyping phases meant spending less time on each of the phases. Had there been more time the phases could have been conducted in a more profound manner.

Visual perception

Visual perception was tested during the iterative design process, and since it was a matter of interpretation, the visual perception could be tested on anyone. However, there was also a deeper level of knowledge involved, which presented itself in features. What features did the operator want, need and request? For the *workshop* group, that question was hard to answer correctly, and it was also where the process was lead slightly astray. Some features that got high points and became focus points were later disregarded by the real end user. Theses features disregarded by the real end user got too much attention in the design process. But, sometimes the real end users does not know which new and *unexplored features* that could be an improvement for them. By talking with others, some features that are considered unconventional may be discovered and discussed.

It was clear that the flaws in the process used in this thesis were not the design but rather the focused on features that the operators were not interested in. Strengthening the answer that to a certain extent it is possible to involve the end user later in the process but also that it is easy to miss important features when substituting the end user from the beginning of the process. The features that were appreciated were the *open box*, the usage of *red/yellow* to display the threat level of the object, in this case, if the object was in or outside a *trigger zone*. The *map* with *history* and trigger zones displayed together and a *toggle option* to show the *distance* to all objects were all appreciated by the operator.

Among the features that through the process of this thesis were voted important, that the operator could see none or little usage for, were the *unknown/alert* symbols, *human/car* classification symbols, the distance indication by box thickness and the mask.

Knowing which features the operators were interested in would have lead the process in a different direction but overall half of the features were appreciated by the end user. Note that the *conclusions* about what the end user likes or dislikes was based on two participants personal thoughts. One of the working as a SO and the other one as a security manager. Had there been more end users able to validate the hi-fi demonstration these results would weight even heavier. If an AO was available for participating. A comparison could have been conducted and discussed.

8.4 Future work

During the design process, several features were discovered, some of these fell outside the scope of this thesis. Nevertheless, the features were equally significant to features covered, therefore they are mentioned below.

Interactive features

Some of the features were dependent on *interactivity* and as such, they are automatically outside the scope of the thesis. But these features are still interesting for future research and in-depth testing. The interactive features are presented below but not further investigated.

A *toggle* function can be thought of as a lamp switch, the lamp is either on or off. As an extension to the function, it can be complemented by a slider, which enables the lamp to be dimmed. The usage of toggles can be a powerful way to remove unnecessary information and as a result highlight interesting information. Several ways of using toggles were thought of.

The usage of toggles and user interaction are important features to incorporate in the future since it directly affects the user experience, however, it is important not to impair the original video since the video also functions as a forensics tool, monitoring past events.

Features outside the video stream window, map, trigger zones and history trail
A requested feature was adding a *map* with a position overview outside the video stream. Both AO and SO wanted the full video stream window to interpret the situation correctly. They also wanted more information about the surroundings. This information were suggested as a feature in the early phases and later mentioned in the validating phase.

To combine a map with the trigger zones for a specific camera and adding a *history trail* and a *distance indication* could be the perfect way of displaying the additional information. Doing this the history trail, distance indication and *trigger zone visualization* would be moved from the video stream, providing a less cluttered way of presenting the object detection.

The radar could provide a precise location for the detected object and a map could provide a way for plotting movements in an easy, non cluttered, manner. The boundaries for the trigger zone would have to be set up by the *installer* and could by doing so be incorporated in the map. A thing to consider was that a map showing movement in real time could be distracting to the operator. A positive or negative consequence depending on the situation.

In general, it is a hard to distinguish *text* on a moving background. Text information could instead be presented outside the video stream. An aspect aired during the lo-fi stage was the suggestion to add the *human classification probability*, a confidence. During the hi-fi validation the classification system was disregarded and opposed to the design principle the end user suggested the usage of text to display the distance to the objects. The information could be of interest in scenarios where it is critical to distinguish the moving object. Another way to use text and numbers, not proposed by the end user, was to add *speed information* for objects. The information could be presented outside of the video stream. Another suggestion, also not from the end user, could be the usage of a counter telling the operator how many identified triggers the system could distinguish in the scene. This could be useful under circumstances where multiple objects are detected in a cluttered scene.

Scalability

When it comes to *scalability* there are several aspects to consider that could affect the product. It is of high importance and benefits that it has a scale potential. Scalability adds new demands and conditions to the product and therefore more aspects have to be considered, more than aspects covered in this thesis.

Scale to multiple monitors When scaling the prototype to *multiple monitors* a need for another solution might be needed. This thesis only covered one monitor and therefore multiple monitors were not one of the aspects during the *iterative design process*. It is possible that it would not be a problem since the end user removed all aspects they though would clutter the scene without adding vital information. They work with several monitors every day and therefore it is likely that their opinions took scaling into account during the testing process.

International scalability For an international company who has customers all over the world, it is important to follow the norms, culture and design guidelines internationally. Further research could be done within this area, as an example, a study of how design standards differs internationally could be performed. Does red always imply alarm, danger? or is red representing something else in other cultures? Since the final prototype does not contain text, aspects such as reading from the right, does not affect the design choose as of now. However, this could be something to consider when developing the whole GUI design where text is included.

9

Conclusion

The purpose of this thesis has been to graphically visualize the combined information that the radar and camera provide together by applying a user-centered design process. The real end user was not involved until the validation at the end. The graphical visualization was iteratively tested by participants with different perspectives with an ambition to mirror the cognitive perception of the end user. The conclusion of this thesis summarizes as follows.

It was possible to practice a user-centered design process with limited access to the end user, to a certain extent. Identifying and defining the real end user was of highest priority, since a test group could not provide a sufficient substitution. Even though substitution is a viable choice, we do not recommend solely using it.

There are several aspects that could increase the likelihood of the graphical visualization fitting the demand of the end users. It is important to listen to the real end users but not solely to them, since that might lead to missing new and unexplored ideas the real end user has not thought about. Results from a wide test group could also indicate which design choices were most likely to live up to the demands of the end user. These aspects together with design principle creates a solid foundation for the graphical visualization design.

Iterative prototype phases theoretically improved the design by enabling testing at different stages of the process. Allowing features to be considered and reconsidered created a refining design process that allowed participants to collaborate and build upon each others ideas. The end result showed that the design process created a perceptually favorable visualization of object detection.

Summarized, the process of UCD with limited access to the end user is a viable way to go but including the end user in the first phase could lead to prioritizing the right features from the beginning. The information can be graphically visualized to fit the demand of the end user and the iterative prototype phases did improve the design process.

Bibliography

Abras, C., D. Maloney-Krichmar, and J. Preece (2004). *User-centered design*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.94.381&rep=rep1&type=pdf.

Adiseshiah, E. G. (2016). *Iterative prototyping to improve the design process*. https://www.justinmind.com/blog/iterative-prototyping-to-improve-the-design-process/.

"Alcser, K., J. Clemens, L. Holland, H. Guyer, and M. Hu (2017). *Cross-cultural survey guidelines, proxy interview*. http://ccsg.isr.umich.edu/index.php/resources/advanced-glossary/proxy-interview.

Boström, M. and T. Claesson (2017). *The title of the work*. MA thesis. Lund Faculty of Engineering, Lund University.

Budge, M. C. and J. S. R. German (2015). *BASIC RADAR ANALYSIS*. Artech House Publishers; London.

Cao, J. (2015). *A guide to prototype fidelity*. http://sixrevisions.com/user-interface/design-mockup-fidelity/.

Communications, A. (2016a). *About axis*. http://www.axis.com/global/en/about-axis.

Communications, A. (2016b). *The axis ecosystem*. http://www.axis.com/global/en/about-axis/the-axis-ecosystem.

Eason, K. (1987). *Information technology and organizational change*. London: Taylor and Francis.

Farrell, S. (2015). *Agile and ucer-centered design*. https://www.nngroup.com/articles/mozilla-paper-prototype/.

Google (2017). *Material design*. https://material.io/guidelines/\#.

Johnson, J. (2010). *Designing with the mind in mind*. Elsvier Inc.

Kumana, D. and J. Dickinson (2014). *Agile and user-centered design*. https://www.thoughtworks.com/insights/blog/agile-and-user-centered-design-0.

Lankow, J., J. Ritchie, and R. Crooks (2012). *Infographics: The Power of Visual Storytelling*. John Wiley & Sons.

Mulholland, B. (2016). *The traffic lights of ux: staying smart with color*. `http://usabilitygeek.com/traffic-lights-ux-smart-color/`.

Nielsen, J. (2009). "Discount usability: 20 years". *Nielsen Norman Group*.

Nielsen, J. (2012). *Thinking aloud: the # 1 usability tool*. `https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/`.

Norman D. A, S. W. D. (1986). *User Centered Systems Design*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Pacheco, M. (2014). *Understanding design fidelity for creating a great product experience*. `https://cantina.co/understanding-design-fidelity-for-creating-a-great-product-experience/`.

Peterson, H. and D. Dugas (1972). *The relative importance of contrast and motion in visual perception.* Rand Corporation in Santa Monica in California.

Ritter, F. E., G. D. Baxter, and E. F. Churchill (2014). *User-Centered Systems Design: A Brief History*. Springer London.

Rubin, J. and D. Chisnell (2008). *Handbook of Usability Testing*. Wiley Publishing.

Smickilas, M. (2012). *The Power of Infographics*.

Spool, J. M. (2004). *The kj-technique: a group process for establishing prioritie*. UIE, User Interface Engineering, `https://articles.uie.com/kj_technique/`.

Standardization, I. O. for (1998). "9241-11:1998, guidance on usability". *International Organization for Standardization*.

Standardization, I. O. for (2009). "9241-210:2009, ergonomics of human system interaction - part 210: human-centered design for interactive systems". *International Organization for Standardization*.

Tullis, T. and B. Albert (2008). *Measuring the user experience*. Elsevier: Morgan Kaufmann.

Ulrich, K. (2003). *Kj diagrams*. `http://opim.wharton.upenn.edu/~ulrich/documents/ulrich-KJdiagrams.pdf`.

Veillard", D. (2017). *Parse4.c*. `http://www.xmlsoft.org/examples/\#parse4.c`. Parse an XML document chunk by chunk to a tree and free it.

Ware, C. (2004). *Information Visualization perception for design*. Elsevier Inc.

Williams, L (1985). *Tunnel vision included by a foveal load manipulation. Human Factors*. Psychology Department at University of South Dakota in Vermillion in South Dakota.

Appendices

Appendix A Mid-fi scenarios

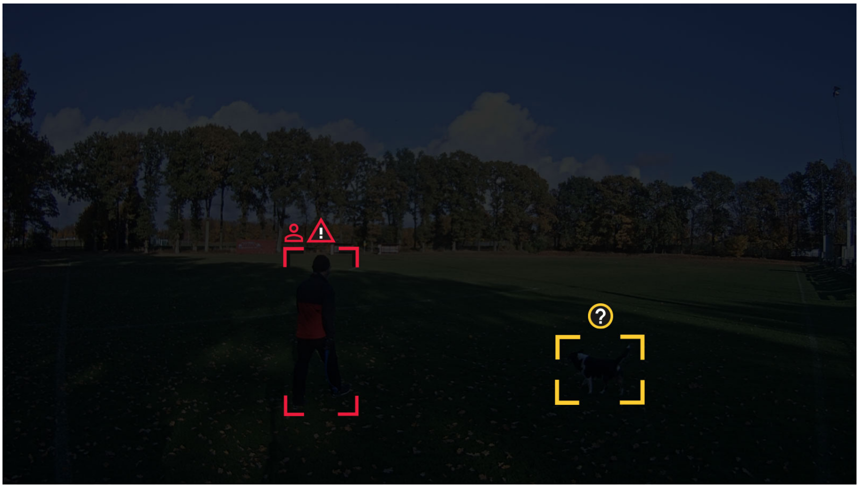

Figure A.1 Scene 2, Combo 7: No mask, Box marking, Id 1

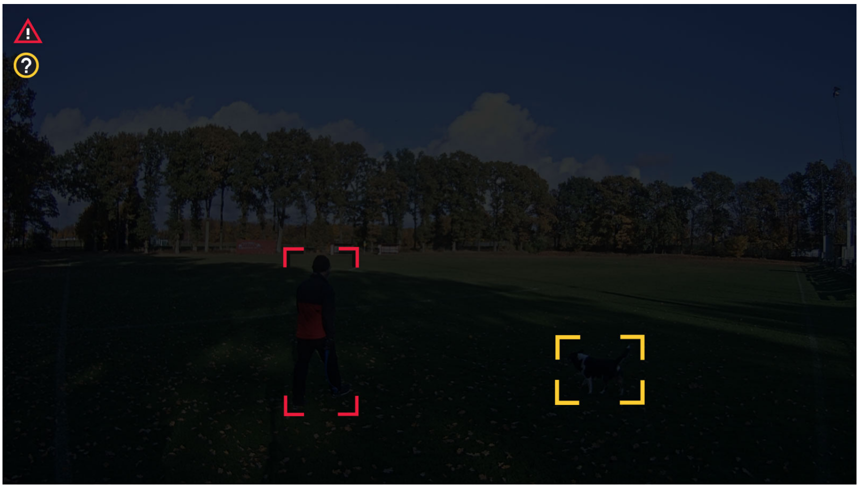

Figure A.2 Scene 2, Combo 8: No mask, Box marking, Id 2

Figure A.3 Scene 2, Combo 11: No mask, Contour marking, Id 2

Figure A.4 Scene 2, Combo 6: Mask, Contour marking, Id 3

Figure A.5 Scene 1, Combo 7: No mask, Box marking, Id 1

Figure A.6 Scene 1, Combo 6: Mask, Contour marking, Id 3

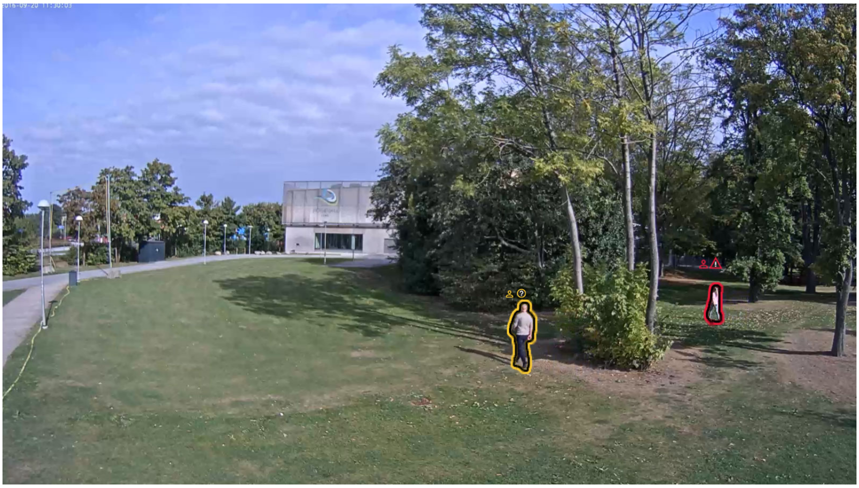

Figure A.7 Scene 1, Combo 4: Mask, Contour marking, Id 1

Figure A.8 Scene 1, Combo 6: Mask, Contour marking, Id 3

Figure A.9 Scene 3, Combo 1: Mask, Box marking, Id 1

Figure A.10 Scene 3, Combo 3: Mask, Box marking, Id 3

Figure A.11 Scene 3, Combo 4: No mask, Contour marking, Id 1

Figure A.12 Scene 3, Combo 6: No mask, Contour marking, Id 2

Appendix B Mid-fi test

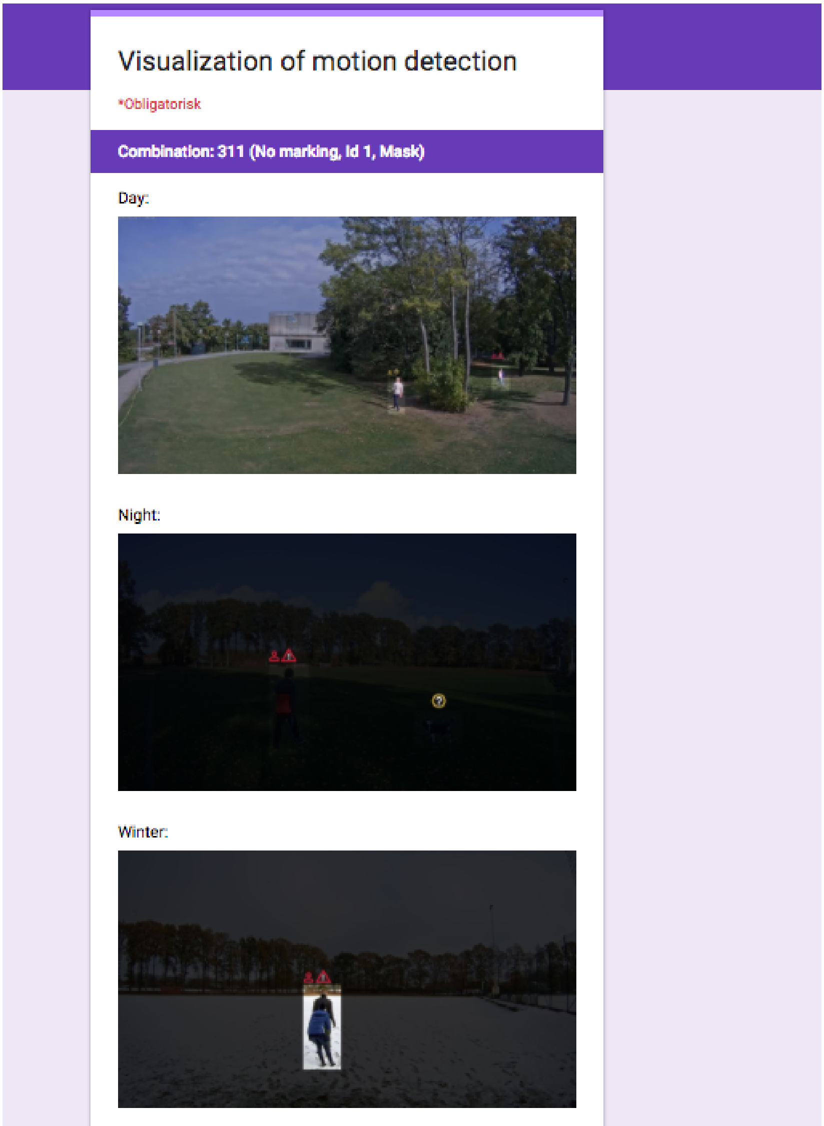

Figure B.1 How the chosen combination was visualized in different scenes.

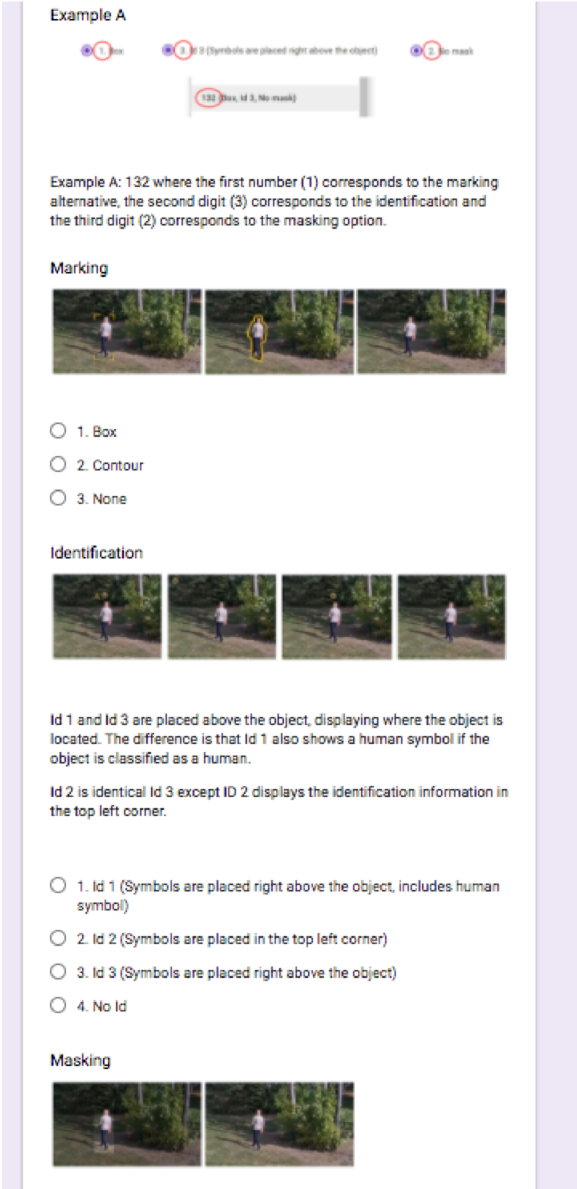

Figure B.2 Part of the view for choosing the different aspects of the combination.

Appendix C - Hi-fi validation plan

Hi-fi validation plan

A validation plan for testing of the high fidelity prototype.

Purpose

Validate the features and visualizations that create the prototype and evaluate if it satisfies and fulfills the needs of the end user.

Background

“Ni kommer att få se konceptuella demos på hur ett eventuellt framtida gränssnitt för motion detection skulle kunna se ut. Det är i nuläget endast på en grov inspiration nivå men för att kunna avgöra vilka delar av konceptet som skulle kunna vara användbara skulle vi vilja ta del av era tankar och funderingar. Ni kommer att få se 7 st demos som speglar samma koncept i olika situationer. Vi kommer att ställa frågor efterhand om konceptet i sin helhet samt för de olika delarna av det.”

Demonstration videos

#1 - Vs6	(röd, utan mask)
#2 - Parking 5	(röd, utan mask)
#3 - Vs23 el. LopX	(röd, gul, utan mask)
#4 - Golf 4	(röd, utan mask)
#5 - Golf 4	(röd, mask)
#6 - Parkering 5	(röd, mask)
#7 - Kart demonstration	(röd, utan mask)

Main focus questions

1. What triggered, where? (Pixel positioning)
2. Illustrate/ highlight the triggered area and trigger objects. (Marking)
    a. Color coding due to threat level
    b. Alert symbols due threat level
    c. Classify objects, visualize objects in the picture. (Identification)
    d. Ability to mask unimportant areas (Masking)
3. Ability to get a geographical overview, a map. On the map visualize motion, and trigger detections. (Geographical positioning)
    a. History path
4. Ranking the needs in regard to the demonstration of the prototype.

Concluded questions

Pre demonstration questions

1. Hur fungerar ert arbete idag? Exempelvis om det sker en objekt detektion, Larm procedur?
    a. Vilka funktioner är bra?
2. Vad ser ni för förbättringspotential inom visualisering av objekt?
    a. Vilka funktioner är efterfrågade?

Post demonstration questions

3. Vad är ert första intryck av scenen? (Allmänt, “on top of the head”)
4. Hur skiljer sig demon från det system ni använder i dag i dina ord?
5. Vad, om några, är de påtagligaste försämringarna jämfört med det system ni använder idag?
6. Vad, om några, är de påtagligaste förbättringarna jämfört med det system ni använder idag?
7. Hur resonerar du kring valet av öppen box som markering?
    a. Är det bättre eller sämre än en sluten box? Varför?
    b. Vilken annan typ av markering hade varit lika bra/ bättre? Varför?
8. Hur resonerar du kring färg mappningen i videorna?
    a. Vad innebär rött?
    b. På vilka sätt kan rött användas?
    c. Vad innebär gult?
    d. På vilka sätt kan gult användas?
9. Hur resonerar du kring identifierings symbolerna (Människa /bil)
    a. Är de av intresse för operatören, eller överflödiga?
    b. På vilka andra sätt kan symbolerna användas?
    c. Finns det andra symboler som är mer av intresse än människa/bil?
10. Hur resonerar du kring hotnivå symbolerna? (Varningstriangel och frågetecken)
    a. Är de av intresse för operatören, eller överflödiga?
    b. Vad kan innebörden av triangeln vara?
    c. Finns det andra användningsområden för triangeln?
    d. Vad kan innebörden av cirkeln vara?
    e. Finns det andra användningsområden för cirkeln?
11. Hur resonerar du kring användandet av en justerbar mörkläggnings mask för att highlighta områden av intresse?
    a. Vad, om något, är negativt med en mask?
    b. Vad, om något, är positivt med en mask?
    c. Vilka användningsområden hade funnits för en mask?
12. Vad är dina spontana tankar kring kartan?
    a. Är det en nackdel eller fördel att kunna rita ut historik där? (on/off funktion)
    b. Var och hur hade du helst fått historiken presenterad för dig?
13. Hur skulle tjockleken på boxen kunna avspegla avståndet?

What data should be collected?

Ifrån “Concluded questions” samlas information i form av subjektiv kvantitativ samt subjektivt kvalitativ data.
Generella frågor skapar en initierande bild.
“Think-aloud” sker under utövningen för att fånga upp direkta tankar och synpunkter.
“Diskussion och debriefing interview” sker efter varje filmvisning med specifika frågor för att fånga upp en samlad bild.

Demonstration	Funktioner	Frågor	Maxtid
#1. Vs6	1. Rött 2. Människa	-	1 min
#2. Parkering 5	1. Rött 2. Bil	• 3, 8 - a,b • 3, 9	1 min
#3. Vs23 eller Lop	1. Gult 2. Rött 3. Triangle 4. Cirkel	• 3, 8 - c, d • 3, 10	1 min
#4. Golf 4 !mask	1. Markering	-	1 min
#5. Golf 4 mask	1. Markering 2. Mask	-	0.5 min
#6. Parkering 5, slider	1. Mask slidereffekt	• 3, 11	0.5 min
#7. Kartdemo	1. Karta 2. History (nämn on/off)	• 3, 12	1 min

Moment	Delmoment	Material	Tid
Initial briefing	Orientera FP med hjälp av bakgrunden.	Bakgrund Frågor: 1 och 2.	5-10 min
Demonstrationer	Genomför demonstration och “Discussion och debriefing interview”	Demonstrationerna Concluded question Frågor: 3-13	6 min
Debriefing + Discussion + Jämförelser	Kompletterande intervjufrågor samt discussion utifrån underlag Demonstrera efter behov	Demonstrationerna Concluded questions Frågor: 4, 5, 6, 7	10-15 min

Sessions, participants

1. Technical development of video and remote services manager
2. Video Surveillance Operator

Test environment/ tools

The tests will be performed in a familiar environment, with video prototypes. Video prototypes demonstrate different features and functions to be evaluated and validated.

Test roles

Test leader - active interactive, interviewing, discussion and debriefing, taking notes
Protocol writer - passive, complementing questions, taking notes
Supervisor - introduction