Summary

This paper describes the design of interfaces for presenting video summaries from drone swarms in the context of Search and Rescue (SAR). The use of drone swarms for SAR missions is a current topic of research, that is motivated by enabling faster search of people in distress in larger areas. Enabling this requires utilising the video feeds that each drone in the swarm can capture. This topic has been a focus of the HERD project, which the research presented in this paper is also a part of. However, previous studies conducted with the Danish Emergency Agency Management (DEMA) has revealed that viewing multiple live video feeds simultaneously for prolonged periods of time with the attentiveness necessary is impossible. Therefore, this paper looks at the concept of utilising Computer Vision to detect objects of interest from the video feeds. These detections can be presented in a summary of detections, allowing for the SAR personnel to examine the potentially important parts of the recorded footage at their own pace, thus reducing the risk of missing any essential details.

Therefore, the purpose of this paper is to examine ways of designing such a summary of drone swarm detections while ensuring that the users avoid experiencing the situation awareness demons, information overload and attentional tunneling. This was done by looking into related work which describes ways of presenting video summaries, the way humans interpret images, and ways of combatting situational awareness demons. Based on this related work we constructed a set of design principles that guided the design of a number of mockups. The mockups were shown to DEMA's head of drone operations in Jutland, Denmark and a professional drone system developer from Robotto. They provided valid feedback that alongside the aforementioned design principles guided the design and development of two functional prototypes. The prototypes both presented a summary of AI detections from a drone swarm as an interactive storyboard, and allowed for user-controlled filtering of the summary with the purpose of minimizing information overload for the users. The prototypes presented the filtering functionality in two different ways, and this constituted the independent variable in a subsequent online user study. This user study had 8 participants, consisting of drone operators, drone system developers and students who had worked with interfaces for drone swarm systems. During the study the participants performed a task using each prototype, where the given scenario was that a SAR mission where a person had gone missing was ongoing. The goal of the task was then to identify a number of detections that showed the missing person and their personal belongings.

The results of the study showed that presenting a summary in the form a story board was effective as when using one of the prototypes the participants managed to correctly identify 4.75 detections out of a possible 6 within just 5 minutes, while only marking 1 detection incorrectly. Furthermore, the participants commented that the filtering functionality aided them in avoiding information overload by limiting the detections presented in the summary to a manageable subset. However, attentional tunneling remained a challenge for participants which calls for further research into this topic.

Interfaces for Presenting Summaries of Detections from Search and Rescue Drone Swarms

ANDREAS DAUGBJERG CHRISTENSEN, Department of Computer Science, Aalborg University, Denmark
SHPEND GJELA, Department of Computer Science, Aalborg University, Denmark

Drones are currently being used in Search and Rescue (SAR) missions enabling ground operators to scan a large area for people in distress by utilising the drone's video feed. Research has begun looking towards drone swarms which would enable the search of an area faster compared to using a single drone. Many aspects of the SAR mission will be automated including path planning, and detecting objects of interest such as people in distress. Using AI to identify possible targets will be essential as it is not possible to have an operator observe live video feeds from multiple drones over a prolonged period of time with the necessary attentiveness. Therefore, the video feeds must be summarized with the possible objects of interest being highlighted through augmentation or annotations to help the operator quickly understand the video feeds recorded by the drone swarm. In this paper we examine how summarization of video feeds from a drone swarm can be used to aid SAR operators during missions. Furthermore, the focus is on designing a user interface that among other features includes filtering functionality that intends to combat the situational awareness (SA) demons, information overload and attentional tunnelling. To do this, we look into theories on how humans interpret and understand images, and how to apply these theories in the context of drone swarms used for SAR missions. These theories revolve around making it easy to identify the key elements in an image and do so quickly. In the project several user interface designs for highlighting the findings of each individual drone and summarizing the findings of the drone swarm as a whole are explored. Some of these designs were presented to the Danish Emergency Management Agency's (DEMA) Head of Drone operations in the Danish region of Jutland, and a developer from Robotto, who work with integrating drones and AI in the context of SAR. Based on their feedback, two prototypes were developed that present a summary of a drone swarm's detections as keyframes in a story board while enabling filtering of the summary based on time and the category of the detected objects. The prototypes were used to conduct an online study with 8 drone operators, drone system developers, and university students as participants. During the study participants were instructed to complete one task using each prototype with an on-going SAR mission given as the imagined scenario. The results showed that presenting a summary as keyframes in a storyboard allowed for participants to correctly identify detections of interest with a significant degree of success. On average the participants correctly marked 4.75 detections out of a possible 6 when using one of the prototypes. The participants also stated that filtering was useful for avoiding information overload.

1 INTRODUCTION

Using drone swarms for SAR is becoming increasingly popular for both researchers and real life emergency agencies. Having multiple drones overfly an area of interest means that an area can be searched quicker than what would be possible with traditional methods, such as searching the area on foot or using a single drone. Though, having multiple drones that are all providing live video feeds to be observed by the SAR personnel leads to another challenge, that is both maintaining an overview of all the incoming stimulus, and noticing any details of importance. The extent of the challenge increases, as the size of the drone swarm increases. Utilising Computer Vision is therefore a potential way of complementing the rescuers' ability to discover objects of interest and take full advantage of a large drone swarm.

Manuscript submitted to ACM

Having artificial intelligence (AI) detect elements of interest in the video feeds from the drones can help the human operators identify and become aware of important details, which might otherwise be missed. However, the operator might still experience information overload if the system in a short amount of time presents multiple notifications alerting to detections made, which again becomes increasingly probable as the number of drones increases. This motivates the option of storing the detections and providing a summary of the recorded video feeds, such that the operator can go through these at their own pace without becoming overwhelmed.
In this paper we develop two working prototypes that summarize the video feeds of a drone swarm with the purpose of providing the SAR drone operator with an overview of the recorded feeds, and alerting them to elements of potential interest. Among other features we focus on researching how filtering functionality can possibly combat the challenge of information overload while also avoiding attentional tunneling. This research is done as part of the HERD project [7] in collaboration with the Danish Emergency Management Agency who provide relevant domain knowledge about the use of drones for SAR missions, and Robotto [20] which is a company developing solutions that use drones and Computer Vision to aid in SAR missions.

Our contributions are:

- Design principles relevant for visually presenting an interactive video summary from a drone swarm in the context of SAR missions.
- Two working prototypes that present an interactive video summary from a drone swarm, with implemented functionality that allows for filtering the summary on time and the category of the detected objects.
- Exploration of ways for enhancing visual elements to make it easier for an operator to understand the message being conveyed by artificial intelligence that is detecting entities of interest such as people, clothing, vehicles etc.

2 RELATED WORK

This section will highlight relevant research that can help stimulate and drive development for summarizing multiple video feeds in a suitable manner for SAR missions. We focus on describing the parts of the research that is concerned with presenting video summaries, and not on algorithms for extracting key parts that tell the story. Furthermore, focus will be put on theories that attempt to explain how humans process information and how they make sense of what they are seeing. These theories can be transferred to the ways humans interpret visual stimuli, and give a set of tools that can improve the method of conveying messages in visual communication.

2.1 Presenting video summaries

Research has examined ways of summarizing video feeds to provide a quick overview of the video content. Different approaches have been investigated.

In [14] they examine the methodology of video synopsis, to efficiently compress and store video footage from drones for later analysis. They develop a system that can detect any abnormal objects in a video, such as a person holding a gun. These objects are then extracted, brought to the foreground, and stitched together in a very condensed video. For instance, they condense videos with a length of 1-9 minutes down to 0.34-2.2 seconds. However, they do not perform any user study to see if it supports an observer in gaining a quick overview and understanding of the video.

In [16] Mei et al. they develop multiple algorithms that summarize a video by extracting keyframes that accurately represent the content of the video. The performance of the algorithms is measured and compared to similar algorithms.

The algorithms' keyframe summaries are compared to human-selected keyframe summaries which act as the ground-truth in their test. With F-scores ranging from 48.2% to 58.5% the algorithms presented in the paper all perform better than the other state-of-the-art algorithms. However, the paper does not describe any specific method for displaying the keyframes in a comprehensive manner, nor do they perform any user study.

In [17] keyframes are also extracted, however, they further identify regions of interest (ROIs) within each keyframe which are used to construct a collage summarizing the content of a video. They conduct a user study where they compare their system, Video Collage, to other video representation tools. The study showed that their way of presenting collages consisting of arbitrarily shaped ROIs was the most visually pleasing representation.

Girgensohn et al. present two approaches to presenting video feed summaries from a set of stationary surveillance cameras [10]. They implement a timeline user interface that shows the chronological order of events detected by the system, with keyframes from the video feeds being attached to the timeline to represent the detected events. Furthermore, they implement a storyboard user interface which is presented as a collection of keyframes that show detected events over time. Both interfaces allow for further investigation of an event by enabling video playback. They do not perform any user study to evaluate the user interfaces in the paper.

In [9] they also propose event boards which is a collection of detected events in temporal order similar to timelines and storyboards, as ways of presenting video summaries from multiple video feeds. They perform a user study, however, it is mostly focused on evaluating the quality of their video summarization technique and not the presentation of it. Therefore, they do not disclose how the participants were presented to the video summaries, and what their opinion on the presentation itself was. Though, their results were promising as they showed that the participants on average found the video summaries up to 89% as informative as the original video.

A system, named CatchLive, which in real-time summarizes live streams by dividing it into sections, and allows for exploring the highlights from those sections in various levels of detail is presented in [26]. The highlights consist of a snapshot, transcription from the video clip being presented as well as highlights from the chat during that part of the stream. Through interviews with frequent viewers of non-summarized live streams they initially gather feedback regarding the challenges that they are faced with when joining a live-stream after it has begun. These include trying to gain an understanding of the previous parts of the stream, while not missing out on information in the current live stream. They perform a user study with three groups of 16-18 participants each having to test the system by watching a live stream about either stocks, cooking or gaming and answering interview questions afterward. Through the interviews they found that the participants experienced that, a timeline helped them get an overview of the live stream, highlights helped them identify important moments in the stream, and the timeline and highlights combined helped them catch up to the current stream with less interruption compared to rewinding. Though they did find that more information is needed to fully understand the previous parts of the live stream, and another user study showed no significant difference in the understanding of the stream for viewers having the summaries provided by CatchLive, compared to another group of viewers that did not.

The generation of textual descriptions that summarize the content of a video is presented in [27, 28]. This allows for a textual description to be associated with a keyframe or short clip in a longer video, which can aid the user in understanding that part in particular as well as the video in its entirety.

In [23] they implement the interactive user interface, VideoForest, that presents a video summary using a tree-like structure with each scene in the video being presented as a branch in the tree. Their interface includes a timeline, and the integration of comments made by previous viewers of the video. The sentiment of these comments is also shown in the interface, to guide users in deciding if they should explore a specific scene further. They conduct an interview with

13 expert participants to evaluate their interface. They find that the features of the interface are helpful for highlighting the eye-catching and evocative parts of the video, and understand the opinions of prior audiences regarding specific scenes.

2.2 Improving visual communication by enhancement

As described above, multiple videos can be summarized through different visual representations. Furthermore, research has also been done in relation to how visual representations can be further enhanced to help convey a message such as ensuring the user is made aware of the important key elements. This can also be seen as improving the summarization of the visual content, which quickly allows the viewer to identify what is of note. To understand what theories and tools could be useful for highlighting elements, it is necessary to investigate how humans interpret images. Research that examines this issue is presented in the following.

2.2.1 Human interpretation of images.

There are different theories relating to humans interpretation of images. These include Gestalt Laws of Perception, Semiotics Theory, and Cognitive Load Theory. Several papers have explored their use and how effective they are. This will be further examined in the remainder of the section.

Gestalt Laws of Perception. The Gestalt Laws of Perception were developed from the theory of Gestalt Psychology. There are several laws, also sometimes referred to as principles, that can be applied to the design of graphical interfaces to make them easily understood and pleasing to look at. In [24] they create a study to compare two of the Gestalt Laws, namely proximity and similarity to see which approach is strongest when it comes to how people consider objects to be part of the same group. In their paper they define similarity as having either the same color or shape. The result of their study reveals that there is a slight favor for a combination of both color and shape against just applying proximity.

In [19] they talk about how the effectiveness of visual communication design can be further improved to ensure the message of a design is properly conveyed to the observer. One of the primary challenges they present are keeping designs aesthetically pleasing but also functionally-legible to ensure the intended message is understood. Color and contrast are two important tools at the disposal of designers for such a task, and they explore how color and contrast in conjunction with the Gestalt Laws can be used to improve the effectiveness of visual communication design. Through the paper, examples are given for several of the Gestalt Laws that emphasizes the usefulness of utilising color and contrast when working with visual communication design.

Visual Semiotics. The theory of semiotics is about how meaning is created and gets communicated and finally interpreted by humans, based on a set of *codes* that they each individually hold. The codes that each individual applies is based on previous experiences and the context which the message that they are trying to decipher appears in. One of the core fundamentals is the *sign*, that is anything used to convey or communicate a meaning. The *sign* is thus said to be comprised of the *signifier*, that is the media or material used to signify a concept referred to as the *signified*. Finally *denotation* is the literal meaning of a sign with *connotation* being the implicit meaning [4, 5, 22]. Visual Semiotics is a subset that focuses more specifically on how images can be used to communicate a message [13].

In [25] they conducted a study in order to investigate how the theory of semiotics can affect poster designs for better or worse. This was achieved by forming two groups, an experimental group and a control group. The experimental group was taught the principles of semiotics and would apply these to their designs. In the end the result exhibited a slight favor towards the experimental group on topics like creativity, aesthetics, typography and overall score. This concludes that the semiotics theory helps strengthen the overall visual tension in images, in this case posters. The

paper also touches on how a designer ensures the message comes across clearly to the observer by playing around with icons and symbolism. Thus messages can be embedded by combining icons and symbolism, like elements being given different textual patterns than their natural occurring pattern. Furthermore, visual marks when designing graphics are important for communicating the intended message - including texts, pictures, images, colors and textures [25, p. 9].

Cognitive Load Theory. This theory relates to the way which humans process information, and the limitations that humans experience during this process. When processing information or performing some task that increases the cognitive load, the working memory is used to maintain an understanding of all the information being processed. However, the working memory's capacity is limited and if this limitation is exceeded it results in a deterioration of learning and limited understanding of the information [12]. Therefore, this theory is also relevant in the field of human-computer interaction and when presenting information to a user [12]. In [15] they propose multiple ways of reducing the user's cognitive load in multimedia learning. For instance, they suggest taking advantage of both the visual and verbal channels as humans can more easily process information simultaneously if it is received through both visual elements and audio. Other ways, of reducing cognitive load when it is at capacity is segmenting and pretraining. Segmenting covers the concept of splitting the presentation of information into multiple segments and allowing for time in between each segment, which enables the user to focus on one task at a time. Pretraining is the idea of providing instructions and training the user prior to them using a system, as this may reduce cognitive load once they use the system. To limit the risk of exceeding the user's cognitive capacity the authors also propose weeding and signaling. Weeding means that there is an emphasis on only displaying essential information, while signalling is a tool that can be used to highlight what is important while still having less essential information presented, should that become necessary.

2.3 Situational awareness in multi-drone systems

Maintaining Situational awareness (SA) is essential for the operator of a multi-drone system in the context of SAR [2]. Having a high level of SA makes it possible for the operator to make better decisions in an ever-changing SAR mission. Situational awareness is a term that describes one's awareness of the information that is necessary for completing the task at hand. The definition of SA is split into three levels [8], with level 1 SA being the *perception* of each relevant element in the environment. This can be limited if the system does not present all the necessary information or if the information is presented in such a manner that makes it difficult for the user to understand it. Level 2 SA is the *comprehension* of how the level 1 elements relate to the task goal in combination with each other, while level 3 SA is the *projection* of the elements' future status.

Designing a system that increases SA can be difficult and certain challenges, known as SA demons, often arise in the development of systems where maintaining the user's SA is critical. Some of these demons include, attentional tunneling, requisite memory trap, information overload, and out-of-the-loop syndrome [8]. In [8] they propose a set of design principles that can be used to address the SA demons. For instance, information overload can be avoided by allowing the user to perform information filtering thus deciding on the content that they want visible. This is reiterated in [21], where *Overview first, zoom and filter, then details-on-demand* is described as a golden rule for designing comprehensible, predictable, and controllable interfaces.

In [2] they collaborate with emergency responders to co-design a multi-drone system for SAR that aims to address the SA demons described in [8], including three additional demons that they have identified through their previous work with Unmanned Aerial Vehicles (UAVs). By getting feedback from the emergency responders they make key observations regarding the SA demons and how to combat them. In regards to attentional tunneling for instance, they

learn that the emergency responders want the ability to focus on a particular detection when the system deems it necessary. In that case they do not see attentional tunneling as a demon but rather a necessity. However, maintaining an overview of the other drones is still important. Another SA demon that they seek to address is information overload. In regards to that, the emergency responders did not experience information overload when using the system designed in the study, however, some elements were distracting which made it clear that only the relevant information in the specific context should always be visible. Additional information should be available on user request. In their work, they conduct 6 co-design sessions to elicit feedback, but they do not test their system in a real-life environment which could have elicited different responses regarding information overload for instance, as that could more easily occur in a stressful situation.

3 PREREQUISITE KNOWLEDGE REGARDING DEMA

As our research is focused on drone swarm systems within the context of SAR, and is done in collaboration with DEMA we will describe their current approach to the use of drones during SAR missions. This is to get a better understanding of how each element of a drone swarm system, such as a video feed summary, could be integrated into their workflow.

The current workflow at DEMA when conducting SAR missions involves a drone operator, an observer and a single DJI drone. An operator will manually take control and fly the drone or alternatively use the built-in functionality of the DJI application to program a flight path for the drone. Meanwhile the observer is monitoring a delayed video feed from the drone on a large monitor which is mounted to the back of their operational van.

Multiple studies in the HERD project have already explored the use of drone swarms for SAR operations [3, 11], such as how to incorporate live video feeds from numerous drones at once and how to display this to the observer [6]. This research has resulted in the HERD system which is currently still in development. This HERD system relies on using a tablet for controlling the drone swarm, an application running on each of the drones, and finally a server for transmitting data between the tablet and drones. An illustration of the HERD system setup can be seen in Figure 1. The research done in this paper will be done with the purpose of highlighting features needed when integrating a video feed summary system into the HERD system.

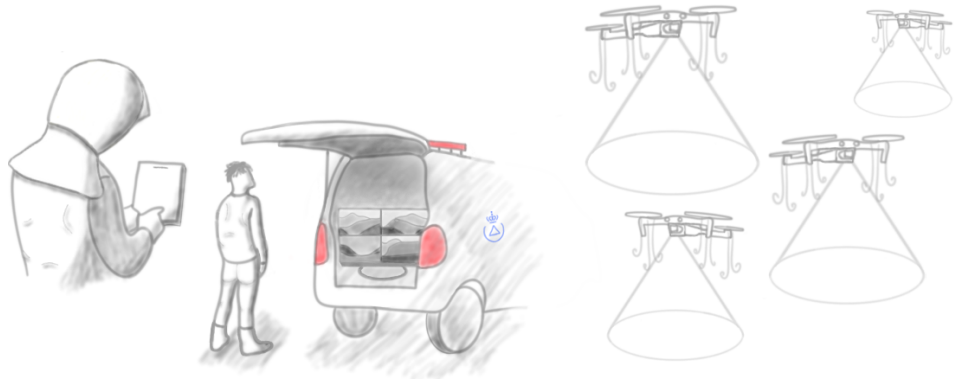


Fig. 1. The current setup of the HERD system with a single operator using the tablet to give instructions and monitor multiple drones in the air at once. The observer is positioned behind the van monitoring the video feeds of the drone swarm on a large screen.

During these studies certain issues and concerns were revealed. For instance, they would often operate in cold weather which would affect their ability to use their fingers to touch a screen for prolonged periods of time. Furthermore, user interfaces with relatively small elements were difficult to interact with due to having large hands, meaning they were concerned about cluttering the screen, thus resulting in little screen space for each element. It also became apparent that the operators have varying levels of expertise when it comes to computers, therefore, systems must be simple, easy to understand and straight forward to use. Finally interviews with DEMA revealed that they were anxious about giving up control and relying entirely on an artificial intelligent driven drone swarm system. They would rather have a system that supports them during their missions, by using Computer Vision for detecting entities of interest while always keeping the human operators in the loop. The research presented in this paper will take into account what has previously been highlighted during interactions with DEMA when developing appropriate mockups and prototypes.

4 RESEARCH PROBLEM

The prerequisite knowledge obtained from previous studies with DEMA regarding the use of drone swarms for SAR missions has highlighted a number of challenges and opportunities within this context.

Interviews conducted with DEMA previously have concluded that observing more than four video feeds simultaneously with the care required for SAR missions would not be possible [6]. Therefore, the system must provide the observer with assistance that aids the observer with noticing the elements of interest. Such assistance could be provided by AI that utilises Computer Vision to detect and notify the user of noteworthy detections made in the video feeds. However, having a large drone swarm could then potentially result in many simultaneous notifications which may overwhelm the observer. Therefore, it could be useful to summarize all the noteworthy detections such that the observer can get a high level overview of the detections and examine each detection in detail when need be.

Furthermore, operators from DEMA have expressed that observing something in the video feeds that may be of interest requires them to alter the flight of the drone to take a second look. This could be avoided if the system automatically stored that potentially important part of the video feed, such that the observer could take a second look while allowing the drone swarm to continue on its designated path.

By looking into the related work it was apparent that summarization of video feeds within different contexts is a topic of interest to researchers. It was evident that many different techniques for presenting summaries have been proposed in the literature. Some of these include storyboards, timelines [9, 10], textual descriptions [28], transcriptions [26], condensed videos [14], collages [17], and keyframes [16].

Additionally, it became apparent that effectively communicating what is noteworthy through visual content is an important aspect of summarizing video feeds. Looking into related work on this topic revealed that different theories regarding human interpretation of images can be applied to communicate messages in visual content effectively. The theories include the Gestalt Laws of perception, Visual Semiotics theory, and Cognitive Load theory.

However, there is a lack of work where the two topics, video summarization and communicating through visual content, are combined to create and evaluate different ways of presenting video summaries through user studies. Especially, in relation to presenting video summaries from drone swarms in the context of SAR missions.

Finally, the work presented in [2] showed that increasing the user's situational awareness is essential when designing a multi-drone system for SAR. They present and address a wide variety of SA demons, providing an overview of the challenges that may arise in the development of such a system. Due to their broad approach however, they leave room for researching how to address specific SA demons such as attentional tunnelling and information overload, more thoroughly. Those SA demons are of particular interest as a summarization system could potentially combat them by

allowing for both maintaining an overview of the drone swarms' findings and each individual drone's findings in more details if necessary.

Therefore, we ask the question: **How can we design and develop a system for summarizing video feeds from drone swarms while addressing the SA demons, attentional tunnelling and information overload, in order to support personnel that are conducting SAR missions?**

5 DESIGN PRINCIPLES

The following section proposes a set of design principles that will help guide the design choices made for the proceeding mockups and subsequent working prototypes. A subset of the principles will originate from previous studies involving DEMA as domain experts. The remaining principles are either inspired by or directly derived from the related work on how summaries can be appropriately presented, how messages can be efficiently conveyed visually, and how the SA demons, information overload and attentional tunneling can be avoided.

Previous work along DEMA highlighted several interesting points worth taking into consideration. These include:

(1) **Minimal interaction** with the handheld tablet due to weather conditions affecting the operators ability to use the tablet for an extended period of time. The path to obtain information should be kept at a minimal, and information should be clear to avoid excessive browsing.
(2) **Large components** as the screen real estate on the tablet is small so enlarging elements to increase readability and to also allow for easier interaction is crucial.

To expand the list of principles further, the related work section highlights numerous valid points that are converted into design principles and listed below.

(3) **Segmenting** is useful to reduce the cognitive load by splitting the information into separate phases that are presented at different times as mentioned in [15].
(4) **Weeding & Signaling** are principles about removing irrelevant information, or highlighting certain elements which require additional attention [15]. Their purpose is to make the content less overwhelming to the user by reducing their cognitive load.
(5) **Grouping items** can as seen in [24] be further strengthened by applying the Gestalt Law of Similarity using a combination of both colors and shapes. This will be particularly useful for constructing visual elements that must be understood as belonging to the same group, or at least be related to each other in one way or another.
(6) **Use of contrast** to alter hue and saturation can attract attention to designated regions of an image as explained in [19]. Attracting the observers attention to the findings on images will decrease the time spent decoding and allow for faster browsing through the summaries but also retain the context of the environment surrounding the findings.
(7) **Utilise colors** to help separate elements from the background and stand out more clearly as described in the Gestalt Law of Figure/Ground. Furthermore, proper use of colors can enhance the Gestalt Law of Good Continuation as explained in [19]. Good continuation will help emphasize and highlight relationships between elements such as a reading direction e.g.
(8) **Global overview** should be provided to ensure that the operator's attention is not tunneled towards a subset of information, while other information that may be of higher importance is not attended to. This allows the operator to have a complete high-level understanding of the situation and make the optimal decision [8].

(9) **Drill down** in the summary for finer level of granularity as seen in [23] with a primary window supplying an overview and a side panel that allows for exploring a specific section in greater detail. The primary window can consist of a timeline with highlights that helps the user understand what has occurred previously [26].

(10) **Reduce display density, but don't sacrifice coherence** is a principle that aims to reduce information overload by spreading out the information on the screen, such that it is easier to process for the user. However, you should be careful with spreading the information so much that it is presented on different pages, as this can reduce coherence in the system [8].

(11) **Represent information timelines** is a useful technique for increasing the user's SA, as it provides data regarding the recording time of the presented information which in turn may influence the user's decision making [8].

6 INITIAL INTERVIEWS

Two initial interviews were conducted with domain experts to give valuable insight and feedback. Prior to these meetings, a set of mockups were designed based on the design principles described in Section 4. These mockups were designed to be an integrated part of the aforementioned HERD system. The first meeting was conducted with Robotto on the 9th of March, 2023. The second meeting was held on the 14th of March, 2023, and representatives from both Robotto, University of Southern Denmark (SDU), Aalborg University (AAU) were present as well as DEMA's Head of Drone Operations in Jutland.

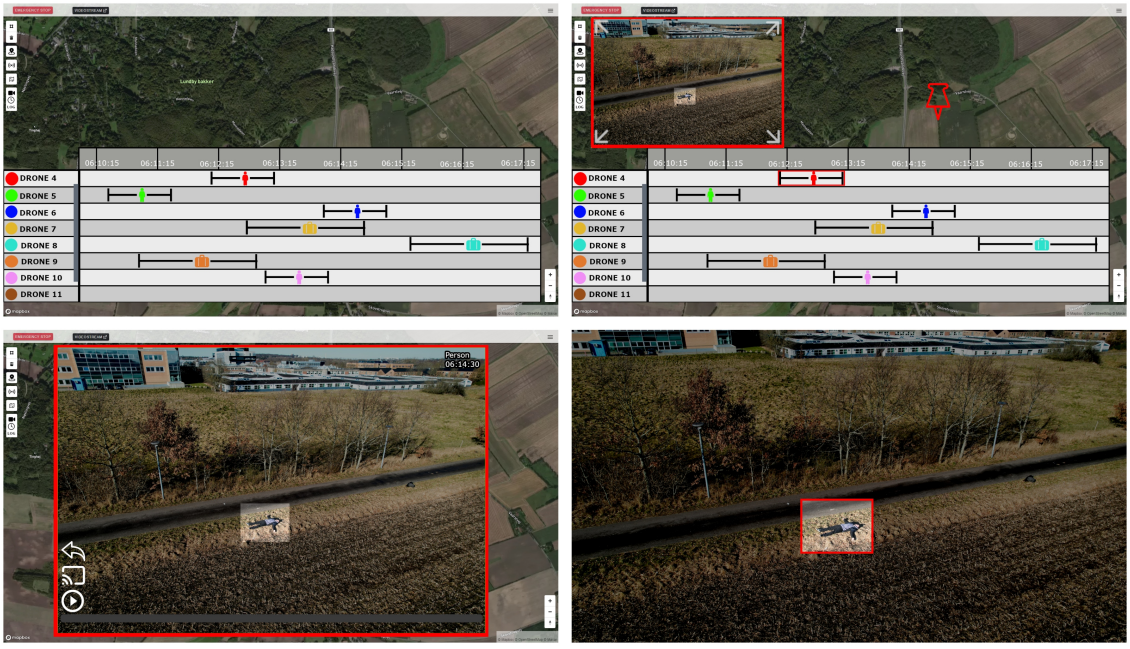


Fig. 2. *Top-left*: The summary presented as a timeline that is overlayed on top of the interactive map. *Top-right*: A preview of the detection is visible after selecting it on the timeline. *Bottom-left*: The preview is enlarged and the corresponding video recording can be played from the tablet or sent straight to the large monitor. *Bottom-right*: An enhanced keyframe with a detection of a person. The saturation of the surrounding environment has been altered and a bounding box encapsulates the detection.

At both meetings the mockups were presented and used as a basis for obtaining knowledge about what is required of a system that summarizes the findings of a drone swarm during a SAR mission. Some of the mockups which were presented can be seen in Figure 2. During the first meeting the set of mockups were printed and handed to a developer from Robotto, while at the second meeting the mockups were displayed with the use of a projector to allow all participants to follow along.

Robotto expressed concerns about showing images in their entirety, as the operators might be using small display screens like handheld controllers or tablets. It is therefore beneficial to crop the images and leave the detection in focus. Robotto were also fond of the general concept of summarizing and presenting detections to the user and thought it was worth exploring as a potential feature. The representative from Robotto also held a demonstration of their own system that lacks any built-in summarization, but instead saved the detections to a folder for post analysis.

In the second meeting it was primarily the representative from DEMA who provided feedback. He was quick to dismiss the need for information about which drone had made a specific detection. He was much more interested in what was detected, together with when and where a detection was made. Furthermore, timestamps showing either the time of detection or the time since the detection is preferred, and a timeline would be suitable for presenting this. Once again, DEMA also emphasized the importance of keeping the interface simple so that anyone can operate it with minimal technical expertise. A total of 12 different ways of enhancing the keyframes by changing HUE, saturation, blurring and colored borders was presented. DEMA was unable to say which of the 12 is the most suitable approach but suggested keeping multiple options as each individual will interpret them differently and for that reason it is best to provide them with numerous options.

7 PROTOTYPE: INTERACTIVE STORYBOARD FOR SUMMARIZATION

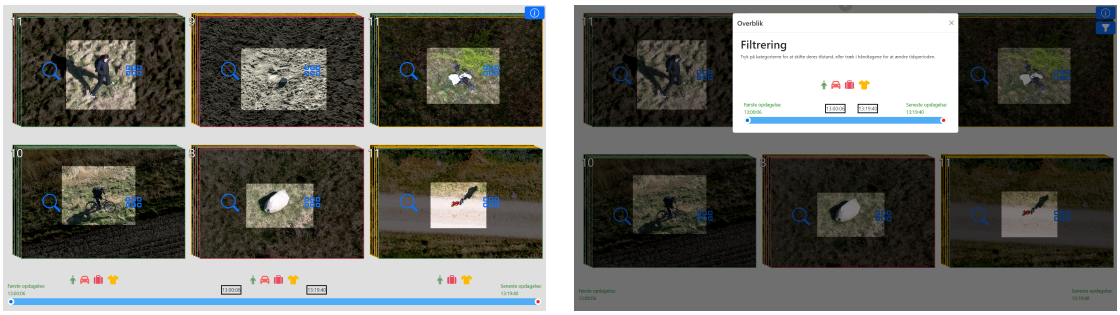
The feedback received from presenting the mockups in the initial interviews was processed and used to proceed with designing and implementing two prototypes. Additionally, the prerequisite knowledge previously obtained from DEMA as part of the HERD project, and the design principles which were derived from the related work were used to shape the design of the prototypes. This section will describe the design and implementation of the prototypes, and the design principles applied to the prototypes. Furthermore, it will describe how each of the prototypes are meant to facilitate a study that examines the user's experience of information overload or attentional tunneling when using a system that summarizes detections from a drone swarm.

7.1 Prototype design

As mentioned two prototypes were developed and both prototypes have large similarities in their designs, with the major differences being the way that functionality for filtering the summary is presented. Therefore, the design choices behind the elements that they have in common will be described first, and the differences between them second. The two prototypes can be seen in Figure 3.

A major design change that was decided upon for both prototypes which differed from the mockups presented during the initial interviews, was to present the summary as a storyboard containing a collection of keyframes rather than a timeline representing each detection as an icon. The reasons for this include DEMA expressing that showing information about which drone in the drone swarm had made a detection, as was the case in the mockups, was unnecessary. This provided the opportunity to adhere to design principle 4, *weeding & signalling*, as the non-essential information was removed. Furthermore, the prototype design presents more information by showing up to six keyframes from separate detections without any user interaction, which follows design principle 1, *minimal interaction*. Comparatively, the

mockup design required the user to click each detection on the timeline for a keyframe from that detection to be presented.



(a) Prototype A: Anchored Filtering.

(b) Prototype B: Hidden Filtering.

Fig. 3. (a) shows the summary in Prototype A with the filtering functionality being available through the interactive icons and timeline in the bottom. (b) shows the filtering functionality page opened on Prototype B, with the summary visible behind it.

7.1.1 *Prototype A: Anchored Filtering.*

Prototype A can be seen in Figure 3a. The detections are shown as layered keyframes referred to as stacks, with each keyframe having a colored border representing the category of object that has been detected on that keyframe. The size of the keyframes is chosen with the purpose of them being large enough to view without interacting with the system, while also presenting as many keyframes as possible. This decision is motivated by design principle 2, *large components* as it eases the comprehension of images. Furthermore, this design is meant to provide a *global overview* and ensure that the user can have a complete understanding of the drone swarm findings during the SAR mission, thus avoiding attentional tunneling.

The colored borders are linked to buttons above the timeline which also represent the category of objects that has been detected in the two stacks placed directly above, and this makes use of design principle 5, *grouping items*.

When there are multiple detections within the same time frame they are stacked on top of each other, with a number in the top-left corner that represents the number of detections in the stack. Each stack can be expanded such that the keyframes in that stack are presented separately. Presenting the summary in this manner is inspired by [10], and follows design principle 3, *segmenting* as it compartmentalizes the detections which should reduce cognitive load.

The timeline shows the time interval in which the detections have been made. It is split into three parts with the detections being placed in a stack that corresponds to the time that they were made. For instance, a detection that was made in the second third of the time interval shown on the timeline is placed in one of the stacks in the middle column. On the timeline there is a range tool, that allows the user to select a time interval. This functionality makes it possible to filter the summary, such that it only shows a subset of detections made in a chosen time interval. The use of a timeline follows design principle 11, *represent information timelines* as it provides the user with an understanding of when the detections were made, which could aid the decision-making process in terms of deciding which keyframes to examine.

As mentioned, there are buttons above the timeline with icons representing the categories of objects that are in the detections in the two stacks above. These can be selected and deselected such that only detections from those user-selected categories are shown. Together, with the option to filter the summary on time this functionality is motivated by [8] which describes how user-controlled filtering can help reduce information overload. It also follows

design principle 9, *drill-down* as it allows the user to focus on specific parts of the summary while being able to easily zoom out and be presented with a greater overview.
Each keyframe can be enlarged in the overlay page which opens when clicking the left side of they keyframe as is indicated by the magnifying glass icon. The overlay page can be seen in Figure 4. The keyframe itself is cropped such that it contains an enlarged view of the detected object while still showing some of the surrounding area for context purposes. Furthermore, the keyframe has been manipulated by changing the exposure of the surrounding area, which in turn highlights the detected object. This manipulation is motivated by the related work described in Section 2.2, and follows design principles *weeding & signalling*, *use contrast*, and *utilise colors* which all aim to steer the user's attention towards the detected object such that the time spent decoding the image is decreased and the user's cognitive load is not increased excessively.

Fig. 4. The overlay page that is shown when clicking the magnifying glass icon. The enhanced keyframe is shown with the timestamp and category for the detection presented above. Additionally, there is a button to mark the detection as interesting. On the right, photos of the detected object from alternative angles captured by the drone swarm are shown. The overlay also serves as a carousel that enables the user to browse through the remaining keyframes in the stack one-by-one by using the arrows on the left and right.

The overlay page includes alternative angles of the chosen detection, a textual description of the category of object that has been detected, and a timestamp. The use of images from alternative angles was inspired by DEMA's current use of Skråfoto which is a service that is provided by the Agency for Data Supply and Infrastructure where aerial images captured from different angles are made available. These latest iteration of Skråfoto images are from 2021 [1], making them somewhat outdated. Therefore, we wanted to investigate the use of images from alternative angles taken immediately after the first image of a detected object is taken. These images were captured by flying around the object and taking pictures from different angles. The user can also add the detection to a list of detections that require further inquiry in the SAR mission by clicking the 'Marker som interssant' (confirm image) button. Finally, the user can click or swipe through the rest of the keyframes in that stack within the overlay page.

7.1.2 *Prototype B: Hidden Filtering.*

The design of Prototype B is very similar to Prototype A, and therefore, the design principles that motivate many of the design choices are the same. The main difference is that in Prototype B the filtering options are removed from the summary page where the stacks of keyframes are presented. Instead they are presented on a separate page that can be opened by clicking the filtering button, located in the top right corner, whose functionality is represented by a funnel icon. Here the user can choose to filter the summary based on time and object category, which is then applied to the presented summary. The object category buttons have the same color as the frames' of the detections from that category, which follows design principle 5, *grouping items*. This can be seen in Figure 3b.

The prototype tests the hypothesis given in design principle 10, *reduce display density, but don't sacrifice coherence*, as the summary page will be less dense because of the filtering functionality no longer being presented there. However, the coherence between the filtering options and the detections that are presented in the summary could become less clear as they are shown on two different pages. The purpose of reducing density is to also reduce information overload, and by presenting the summary and the filtering on two different pages the design principle, *segmenting* is also followed as applying the filtering to the summary and examining the summary is split into two distinct phases. As a consequence of moving the filtering from the summary, more screen space is created for the keyframes themselves, allowing them to become slightly bigger thus presumably making it easier to decipher compared to Prototype A. This attempts to follow the design principle 2, *large components*, though the difference in sizes between the two prototypes is subtle.

The two prototypes are meant to facilitate a study which examines how the user interface for a summary of detections from a drone swarm should be presented, such that a high level of situational awareness is achieved by the user. Furthermore, there is a focus on how the filtering functionality can combat the user's experience of information overload by allowing them to limit the number of detections presented based on relevant filters, such as time and object category. However, the filtering should not result in the user losing awareness of the overview provided by the summary, thus experiencing attentional tunneling.

8 USER STUDY

In order to answer our research question an online study was conducted with drone operators from DEMA, drone system developers from Robotto, and students who work with interfaces for drone swarm systems. The aim of the study was to investigate the overall usefulness of a summary of drone swarm detections, and whether specific features such as enhanced keyframes and filtering functionality are helpful for countering the SA demons, information overload and attentional tunneling. To answer these questions the participants were instructed to perform the task of identifying specific detections using each of the prototypes, and evaluating their experience immediately after each task by responding to a NASA TLX questionnaire [18]. Besides the questionnaire we tracked their performance during the task, by tracking the completion time, correctly and incorrectly identified detections, and interactions with the filtering functionality. Finally, the participants had to answer a set of follow-up questions which again aimed to elicit answers to the research question.

The purpose of the study is to evaluate whether the presented prototypes help the participant gain a high level of situational awareness, such that they can make decisions about how to proceed with the SAR mission. The focus is on the manner in which user-controlled information filtering is made available in each prototype, and how that relates to the user experiencing information overload or attentional tunneling. The independent variable in this study is the prototype that is different for each of the two tasks. As described in Section 7 the main difference between the

prototypes is how the filtering functionality is presented. We will analyze how the change in prototype affects the performance and overall experience of the participants when completing the tasks.

8.1 Participants

A remote asynchronous online study was performed with 8 participants. All participants are licensed UAV pilots. The participants had different occupations as 2 participants were drone operators from DEMA, 1 was a professional drone system developer, and 5 were university students who have all worked with user interfaces for drone swarm systems. A detailed summary of the participants' details can be observed in Table 3. The standard deviation reveals a wide spectrum of experience in the set of participants, and that the answers came from primarily males. The difference in experience can be explained by the various roles of the participating people, as not all of them handle drones on a near daily basis.

Detail	Average	SD
Age:	30.13	6.51
Years of drone experience:	3.25	3.93
Years at DEMA:	4.13	8.07

Table 1. Statistics of the particpants.

Gender	Amount
Male:	6
Female:	2
Other:	0

Table 2. Distribution of gender.

Table 3. Details on the particpants' age, experience and gender.

The online study was conducted through a website that we developed, which guided the participant through a number of steps such that they had sufficient information to perform two tasks using the two prototypes, and finally consider some follow-up statements. The website was in Danish as the participants were mostly native Danish speakers.

8.2 Study introduction

The participants were initially given an introduction to the study, highlighting that the overall purpose of the study was to examine how a system summarizing detections from drone swarms used in the context of SAR missions should look like. They were also asked to accept a consent form which allowed us to collect data during the study for post-analysis. To increase the chances of receiving valid feedback from the study, we wanted to make the environment realistic by making the participants understand that this system would be used during a SAR mission. This was achieved by creating a context video showing a drone swarm taking off, a search pattern being defined, and finally the drones searching the area while recording and detecting objects of interest. The participants were then informed that the drone swarms detections were sent to a summary which they would have to interact with to identify detections which required further investigation.

8.3 Task

The participants were shown a video and a series of screenshots explaining the implemented functionality and instructing them how to use each of the prototypes. To ensure, that the participants becoming familiar with such a system by using one prototype first would not positively bias the results when using the second prototype, half of the participants used Prototype A first and Prototype B afterwards, and vice versa for the second half of the participants.

After the instructional video, they were given the opportunity to try the prototype for 1 minute to become familiar with the functionality. During this practice session the summary was made up of detections that were different from the ones used in the actual task. The scenario and tasks varied for each prototype, and the description for Prototype A was as follows:

"A person is missing, and a SAR mission has been launched with the purpose of finding the missing person or any personal belongings that may lead to the person."

The following detailed information was provided about the missing person:

- 23 year old male.
- 192 cm tall.
- Brown hair.
- Clothing:
  - Black pants.
  - Blue and red-striped t-shirt.
  - Blue and white-striped denim jacket.
  - Black shoes.
- Carrying a black backpack.
- Riding a black bicycle.
- The person was last seen at 12:40 leaving their home in a depressed state of mind.
- A drone operator who was watching live video feeds from the drones, noticed something at 13:10 which may be of interest.

The participants were instructed to identify the detections that matched for any of the items in the summary which showed the person himself, each of his clothing items, or his bicycle. The participant should find these detections during the upcoming task and confirm them as being of interest. The task lasted 5 minutes and automatically ended once the time limit was reached. A message would appear letting the participant know when there was 1 minute left. The summary contained 62 separate detections, containing either people, clothing items, vehicles, or bags. These detections were manually labeled to simulate detections made by an AI model. Furthermore, there were a few detections that were false positives, meaning they were labeled with a category, however, they did not contain any object of interest. These false positives were included to make the set of detections more realistic as AI models are not 100% accurate.

This meant that there was a set of 6 detections that were defined as being of interest by the task description, and the participant had to identify each of these for the task to be deemed complete.

Once the task using one prototype was completed, the participant was introduced to the other prototype in exactly the same way as the first. This means a slideshow of screenshots and a video was shown explaining the functionality and how to use it. They also got to practice using the prototype to further understand the functionality and get accustomed to using it.

In the same fashion as before a task was to be completed using the prototype by the participant. The tasks were structurally identical for both prototypes. However, for each task the participants were instructed to identify a different missing person with completely different items of clothing and vehicle associated.

Immediately after trying one of the prototypes the participants were presented with a NASA TLX questionnaire that they had to provide answers to before continuing. The purpose of this was to record their experience after using each prototype and getting data regarding their immediate impression. The participants were asked the following set of questions from the NASA TLX questionnaire:

- How much mental and perceptual activity was required?
- How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred?
- How successful do you think you were in accomplishing the goals of the task set by the experimenter?
- How hard did you have to work (mentally) to accomplish your level of performance?
- How insecure, discouraged, irritated, and stressed did you feel during the task?

The participants were asked to answer each question on a scale ranging from 1 (low) to 10 (high).

8.4 Follow-up statements

To further gather feedback from the domain experts they were asked to consider a set of statements after having tried both prototypes. The participants could give their opinion using a five-point Likert scale ranging from *strongly disagree* to *strongly agree*. The set of statements can be seen in Table 4. Furthermore, a text box was included so the participant could provide their own comments to each statement regarding their experience and opinions of the prototypes.

Follow-up statements			
1.	The system attempts to sum up the drones' detections, by presenting a keyframe from each detection in a stack across a timeline. The number of detections that are presented at once is however overwhelming.	**7.**	Filtering on Prototype A was intuitive.
2.	The option to filter on time and category helped with showing the detections that you wanted.	**8.**	Filtering on Prototype B was intuitive.
3.	Filtering was useful for avoiding too much information being shown at once.	**9.**	Filtering on Prototype A was easy to access.
4.	You experienced becoming so focused on a specific element that you lost awareness of the remaining elements.	**10.**	Filtering on Prototype B was easy to access.
5.	The system helped you decide which detections to examine before others.	**11.**	Prototype A had the filtering options presented on the same page as the detections. This made it easier to maintain an overview when you could filter and see the result of filtering on the same page.
6.	The information that was presented with each detection, was useful in terms of understanding what had been detected, and deciding if it should be investigated further.	**12.**	Prototype B had a separate page for filtering the detections, and thus had fewer elements on the page showing the detections. This made the page showing the detections easier to maintain an overview of.

Table 4. The set of follow-up statements that the participants provided their opinion on through a five-point Likert scale. For each statement they had the option to elaborate if they desired.

Questions were asked to get in depth feedback from the domain experts and to evaluate their experience of the SA demons, information overload and attentional tunnelling. Furthermore, questions were asked regarding the features implemented in the prototypes, and their preferences between the two presented prototypes.

9 RESULTS

With 8 participants doing the online study, a data set was collected for further analysis. We analyzed the performance related data recorded during the tasks, their answers to the NASA TLX questionnaire, and their subjective answers to a set of follow-up statements. These results will be described in this section, with the purpose of answering our research question and investigating the hypothesis that a system that presents a summary of drone swarm detections while allowing for filtering that summary, combats the SA demons information overload and attentional tunneling. The questions presented in the study and the answers provided by the participants were in Danish, but are translated to English when described in this section.

9.1 Task performance

During each task the system recorded a number of different metrics relating to their performance and their interaction with the summary functionality. A detailed summary of the participants' performance and behaviour when using each prototype can be seen in Table 5. The table shows the mean values, standard deviation, difference between mean values, as well as the t- and p-values which were obtained by running paired samples T-test to compare the data logged during the use of both prototypes. The p-values are compared to a significance level of 0.05, when determining if there is a significant statistical difference between their performance across both prototypes.

Performance metric	Mean	SD	Difference	t	p
Correct markings using Prototype A:	3.38	1.22	-1.38	-2.022	0.082
Correct markings using Prototype B:	4.75	1.20			
Wrong markings using Prototype A:	5	2.74	4	3.433	0.010
Wrong markings using Prototype B:	1	1.00			
Right to wrong ratio using Prototype A:	46.49%	21.62	-36.19	-3.183	0.015
Right to wrong ratio using Prototype B:	82.68%	17.71			
Time to mark a detection using Prototype A:	48.79	19.32	-3.39	-0.292	0.778
Time to mark a detection using Prototype B:	52.17	20.66			
Filter interactions using Prototype A:	17.88	12.16	3.88	1.330	0.220
Filter interactions using Prototype B:	14.00	11.30			

Table 5. A detailed summary of the participants' performance showing the mean values, standard deviation, difference between mean values, as well as the t- and p-values which were obtained by running paired samples T-test to compare the data logged during the use of both prototypes.

For each task the number of detections correctly marked as interesting by each participant was measured, to see if presenting the detections in a summary as is done in the two prototypes actually allows for identifying detections of importance. For each task 6 different detections had to be marked as interesting. When using Prototype A the

participants correctly marked 3.38 (SD=1.22) detections on average, while the participants correctly marked 4.75 (SD=1.20) detections on average when completing the task using Prototype B. When using Prototype A the participants marked 5 (SD=2.74) detections wrongly on average, and when using Prototype B the participants on average only marked 1 (SD=1) detection wrongly. The paired sampled T-test shows a significant difference between the number of wrongly marked detections. 0 participants completed the task using Prototype A, meaning no one correctly marked all 6 detections of importance as interesting. While 3 participants (37.5%) completed the task when using Prototype B.

To see if the summary in either of the prototypes aided the participants with identifying the correct detections while minimizing the number of detections incorrectly marked as interesting, we calculated the percentage of correctly identified detections out of all marked detections. Out of all the detections marked as interesting, 46.49% of them were marked correctly when using Prototype A, and 82.68% of detections marked were correct when using Prototype B. This is a significant difference showing that participants made less errors while simultaneously marking more detections correctly when using Prototype B.

To evaluate whether the filtering is useful for identifying important detections, the number of interactions that each participant made with the filtering functionality was recorded. In Figure 5, a scatter plot is seen which visualizes the relationship between the number of filtering interactions and correctly marked detections in order to reveal any correlation between the two variables. The scatter plot does not seem to show any correlation between the number of filter interactions and correctly marked detections. Additionally, Table 5 shows that the number of filter interactions were quite similar between both prototypes with there being no significant statistical difference.

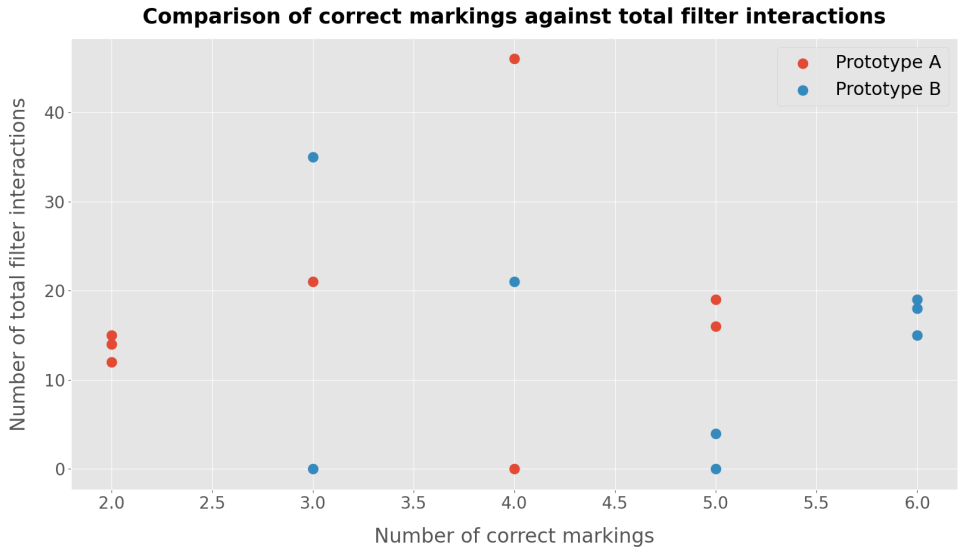


Fig. 5. Scatter plot showing the relationship between the number of filter interactions and correctly marked detections for both prototypes.

When completing the tasks the participants were instructed to identify the correct keyframes that were described in the task description as quickly as possible. Furthermore, Figure 6 shows the mean time taken by participants to mark each correct detection using both prototypes. The bar chart shows that the participants on average spent slightly more time to identify the correct detections when using Prototype B compared to Prototype A. This is also seen in Table 5 as participants on average spent 48.79 seconds to correctly mark a detection using Prototype A, and 52.17 seconds to correctly mark a detection using Prototype B. Furthermore, the bar chart shows that when using Prototype B, participants spent less time to correctly identify detections 4 to 6 compared to detections 1 through 3. Similarly, the time spent identifying detection 3 to 5 using Prototype A was slightly less than the time spent identifying the first couple of detections for Prototype A.

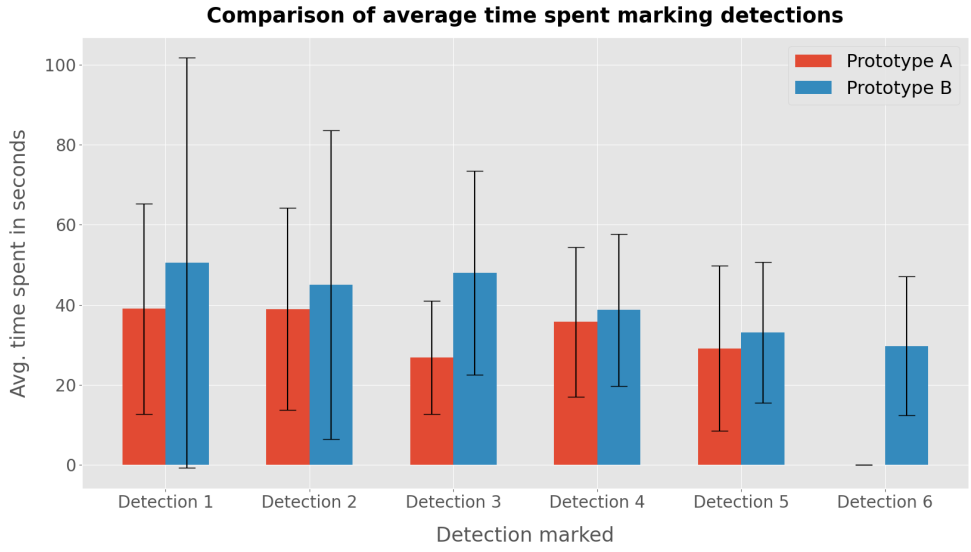


Fig. 6. Bar chart showing the mean time taken by participants to identify a correct detection during the tasks when using each prototype. The error bars are used to represent the standard deviation for each bar.

9.2 NASA TLX Questionnaire

After each task the participants filled out a NASA TLX questionnaire with values ranging from 1 (low) to 10 (high). The average score of the responses for each prototype can be seen in Figure 7. The bar chart shows that using Prototype B was slightly more mentally and temporally demanding while also requiring more effort and causing more frustration. However, the average scores are relatively close across both prototypes.

The participants' average scores for these questions are quite high. For instance, the average score for mental demand when using Prototype A being 6.45 and 7.33 for Prototype B, and the average rating for the level of effort is 6.22 for Prototype A, and 6.89 for prototype B.

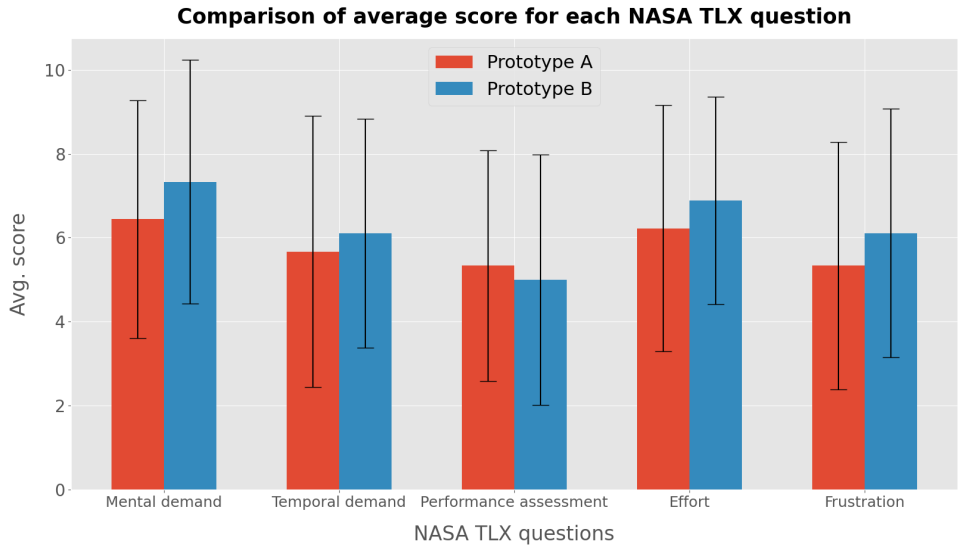


Fig. 7. Bar chart showing the average score from the responses to the NASA TLX questionnaire. The error bars are used to represent the standard deviation for each bar.

9.3 Post-task evaluation

After the participants had finished using both prototypes they were presented with 12 follow-up statements which were to be answered through a 5-point Likert scale ranging from *strongly disagree* to *strongly agree*. The statements can be seen in Table 4. Additionally, the participant could elaborate their answer to each statement by writing in the accompanying text box. Some of the questions were related to both prototypes while others were about one prototype specifically. A selected number of the answers provided by the participants are described in this section.

9.3.1 Filtering.

Statements 2 and 3 were concerned with the usefulness of filtering for limiting the amount of information shown, and avoiding information overload. For statement 2, 75% of the participants agreed that filtering helped with showing the desired detections. One participant elaborated and wrote,

"[filtering on] category helped, but time did not seem that useful given the short timeframe."

and another participant stated,

"I did not use the filtering during the first task, but did so during the second task which helped a lot with structuring my search, and therefore I experienced that I was faster.".

In response to statement 3, 62.5% of the participants either agreed or strongly agreed, while 37.5% disagreed. For instance, one participant wrote,

"Filtering on time was great for starting the search and focusing on the time span that was given in the task description. But hereafter it was difficult to keep track of how the stacks were changing when I moved the timeline".

Similarly, statements 7 and 8 were about the intuitiveness of using filtering on Prototype A, and Prototype B, respectively. Here, 50% agreed, 37.5% strongly agreed and no one disagreed that filtering on Prototype A was intuitive. Though, the elaborated answers did reveal some confusion among the participants with filtering on Prototype A in particular. One participant wrote,

"Filtering in Prototype A was much better ... However, it was very confusing that if I clicked on a bag icon in one [column], then the others were affected as well.".

In regards to the filtering on Prototype B being intuitive, 62.5% agreed while 25% either disagreed or strongly disagreed. The responses revealed that the filtering itself was intuitive but the manner in which it was to be accessed was not. For instance, a participant stated that,

"Having to navigate back and forth to filter was not very intuitive and created more frustration than benefit. However, it was great that there was only one icon per category, contrary to Prototype A.",

which also related to statement 9 and 10, which were about how easy it was to access the filtering in each prototype. Here, 87.5% of participants either agreed or strongly agreed that filtering was easy to access in Prototype A. On the other hand, 62.5% either disagreed or strongly disagreed that accessing filtering in Prototype B was easy. One participant wrote,

"... it was pretty cumbersome to access filtering through a button.".

In relation to the intuitiveness and accessibility of the filtering, participants gave their opinion on statement 11 which stated that, having the filtering on the same page as the detections as was the case with Prototype A made it easier to maintain an overview. Over half of the participants either agreed or strongly agreed with this, while the remaining three were either neutral or disagreed. In response to the statement one participant said,

"It was easier to get an overview, and you could easily change what you were looking for.".

On the other hand, statement 12 postulated that, having the filtering available on a separate page as in Prototype B made it easier to maintain an overview of the detections. On this matter, the opinions were split with 62.5% either disagreeing or strongly disagreeing and 37.5% agreeing. For instance, one participant wrote,

"It made it more confusing, as I had a better understanding of what I was filtering and the result of it when using Prototype A.".

Though another participant agreed with the statement, and wrote,

"I agree. Even though it takes more clicks it seemed more simple.".

9.3.2 Maintaining an overview.

In regards to statement 4, which concerned whether participants became so focused on one element that they lost awareness of the remaining elements, 75% either agreed or strongly agreed that they experienced this. One participant felt that diving into a stack resulted in losing awareness about the other stacks of detections. They wrote,

"When you are within a stack you can easily get tunnel vision and forget that there are other stacks as well."

Additionally, several participants commented that it was difficult to remember what they were supposed to identify during the task. The reason being that the task description was not visible for the participant while they were looking

through the detections. Instead they had to open the task description which would appear on a separate overlay page, and then switch back to the summary to then find the items given in the task description. One participant wrote,

"If you could have the task description open while you're looking at the detections then you would not have to remember as much.".

9.3.3 Detection information.

Finally, statement 6 was about the information that was presented alongside each detection on the overlay page, and whether it was useful for understanding what had been detected and whether the detection should should be investigated further. Here, 62.5% of the participants agreed, with several of them mentioning the photos from alternative angles as being useful in particular. One commented,

"I like the photos from alternative angles a lot, they provide a better understanding of what has been spotted.",

and another wrote,

"Photos from alternative angles were the most useful.".

10 DISCUSSION

The online study that was conducted revealed several interesting findings with some being expected and others being more surprising. These findings and how they relate to the research question we are attempting to answer will be discussed in this section.

10.1 Designing a system for summarizing video feeds from a drone swarm

The overall purpose of the research in this paper was to elicit valid knowledge about the manner in which a system for summarizing video feeds from a drone swarm should be designed and developed. This part of the research question was partially answered through the development of the two prototypes which were used in the subsequent study. It was apparent that presenting detections as a set of enhanced keyframes in an interactive storyboard, allowed for users to identify specific detections containing people and items which were described in a given task description. On average the participants correctly identified 4.75 detections out of a possible 6 in only 5 minutes, when using Prototype B. This confirms that the techniques such as storyboards, timelines and keyframes that are described in [9, 10, 17] as ways of presenting video summaries, are indeed useful and effective for that purpose. When looking at the data gathered from the use of each prototype, it was seen that participants were able to correctly identify more detections with less errors using Prototype B, compared to Prototype A. One potential reason for this could be that Prototype B had the filtering functionality on a separate page, meaning fewer elements on the summary page which in turn could mean the user being less overwhelmed. However, somewhat surprisingly several participants commented that they preferred the design of Prototype A, as it was frustrating to switch between a filtering page and a summary page with the detections. These comments seem to reinforce the two design principles 1 and 10, which are *minimal interaction*, and *reduce display density, but don't sacrifice coherence* (described in [8]) respectively, as the participants want to minimize the amount of clicking required, as well as maintaining the coherence that is achieved when the summary of detections and the filtering are on the same page. However, with the participants preferring the design of Prototype A while performing better when using Prototype B, it could be interesting to investigate the design of filtering in such a system further in a future study.

10.2 Designing to avoid information overload and attentional tunneling

The participants spent around the same time identifying the correct detections on both prototypes. Though overall, they became faster at identifying the correct detections throughout the task. This could be because of participants getting more familiar with the system during the task. However, an explanation could also be that the participants would look through the detections and determine if they were of interest or not. So towards the end of the task the participant would know which detections were definitely not of interest, leaving a smaller subset of detections that could potentially contain an item described in the task description. This relates to a comment provided by a participant who suggested that the system should allow the user to mark a detection as not of interest and subsequently remove that detection from the user's view completely. This could also make the amount of information less overwhelming as the user proceed with the task.

This also ties into the second part of the research question which was concerned with designing the system in a manner which would avoid the user experiencing the SA demons information overload and attentional tunneling. To avoid information overload, the design principle *drill-down* was followed by implementing filtering which would allow the user to control the information being presented. The study revealed that this was achieved to some extent, as participants found filtering useful as it helped them avoid having too much information shown at once. This confirms the guidelines for design presented in both [8] and [21]. Furthermore, we implemented drill-down functionality in similar ways to the techniques described in [23, 26], with positive results which further validates these techniques as effective. Though, we expected to see some correlation between the number of filter interactions and the number of correctly identified detections, however, this was not the case. A reason for this could be that the detections spanned across a relatively short time span, and the total number of detections in the summary was 62. Given this, filtering might have been less useful compared to a summary which presented a lot more detections over a larger time span. This could also be interesting to examine further in a future study. Finally, the aim was also to find ways of designing the system such that the participant would avoid experiencing attentional tunneling, thus losing awareness of the bigger picture. This was not quite achieved in the design of the prototypes as participants did experience becoming so focused on one stack of keyframes that they forgot about the rest of the summary. This confirms the challenge described in [2], because the ability to focus on one detection completely is essential in the context of SAR, however ensuring that focus is not directed to one detection in an excessive manner is difficult. Therefore, it could be interesting to investigate this challenge further, by experimenting with other designs such as a split-screen view showing both a detection in detail and an overview of the summary or the use of notifications to ensure that the user does not forget to consider some parts of the available information.

10.3 Limitations

The research presented has a number of limitations. For instance, the detections that were used in the prototypes were manually captured and labelled to simulate an AI recognizing objects in a video feed. This makes the detections presented in our study less realistic, as a real-life system would use an autonomous drone with AI capabilities to detect objects of interest which could be less accurate and capture the detections differently.

As for the study itself we only had 8 participants and only 2 of them were drone operators which is who such a system in envisioned for. This might limit the validity of the results gathered, even though the remaining participants did have experience with developing drone swarm systems. Finally, the system was envisioned to be used outside during an actual SAR mission. However, the setup in the study was quite different from this with the online study being

conducted inside using a computer. Having a more realistic setup, by doing a field study for instance could perhaps have elicited different and more realistic feedback.

11 CONCLUSION

In this paper we investigated how a system that presents a summary of detections made by a drone swarm in the context of Search and Sescue should be designed. To examine this question we looked into related work, and prior knowledge obtained through the HERD project to create a set of design principles that guided the design of subsequent mockups and eventually two prototypes. The mockups were presented to domain experts in the form of DEMA's head of drone operations in Jutland and a professional drone system developer. This resulted in valid feedback which was used along the design principles in the development of two functional prototypes, which both presented a summary of detections as keyframes in a storyboard across a timeline, while allowing for user-controlled filtering of the summary. These prototypes facilitated an online study that was conducted with 8 participants consisting of drone operators from DEMA, professional drone system developers, and students that work with interfaces for drone swarm systems. Participants had to complete a task of identifying detections which showed a missing person and their personal belongings by using each prototype. The results showed that presenting a summary of detections as done in the prototypes allowed for participants to identify specific detections given to them in a task description. On average participants correctly identified 4.75 detections out of a possible 6, in just 5 minutes when using Prototype B.

Additionally, the purpose of the research done was to discover ways of designing such a system while avoiding the situational awareness demons, information overload and attentional tunneling. To avoid information overload, user-controlled filtering functionality was implemented. Participants commented that they found it useful, and it allowed them to limit the amount of information to a non-overwhelming amount at a time. However, the logged data did not show any correlation between the number of filter interactions and correctly identified detections. Furthermore, the study showed that attentional tunneling was experienced by the participants when using the prototypes. Finally, this research could be expanded upon by conducting a field study, to gain feedback from using such a system in a more realistic scenario.

ACKNOWLEDGMENTS

A special thanks to the employees at the Danish Emergency Management Agency and Robotto for taking the time to provide valuable feedback that helped shape the design of both prototypes. We would also like to thank everyone who actively participated in the study which yielded data for further analysis in order to evaluate the prototypes.

REFERENCES

[1] Agency for Data Supply and Infrastructure. 2021. Information om Skråfoto. https://skraafoto.dataforsyningen.dk/info.html

[2] Ankit Agrawal, Sophia J. Abraham, Benjamin Burger, Chichi Christine, Luke Fraser, John M. Hoeksema, Sarah Hwang, Elizabeth Travnik, Shreya Kumar, Walter Scheirer, Jane Cleland-Huang, Michael Vierhauser, Ryan Bauer, and Steve Cox. 2020. The Next Generation of Human-Drone Partnerships: Co-Designing an Emergency Response System. *Conference on Human Factors in Computing Systems - Proceedings* (4 2020). https://doi.org/10.1145/3313831.3376825

[3] Andreh Bassam Bahodi, Denmark Maria-Theresa Oanh Hoang, and Denmark Rasmus Skov Buchholdt. 2022. User Interface Design for UAV Swarms in Search and Rescue. (2022). https://www.brs.dk/da/

[4] Steven Bradley. 2016. An Introduction To Semiotics — Signifier And Signified. https://vanseodesign.com/web-design/semiotics-signifier-signified/

[5] Steven Bradley. 2016. Denotation And Connotation - Literal And Implied Meaning. https://vanseodesign.com/web-design/denotation-connotation/

[6] Andreas Daugbjerg Christensen, Andreas Skjoldgaard Andersen, Philip Michaelsen, and Shpend Gjela. 2022. Interfaces for Live Video-streams from Search and Rescue Drone Swarms. (12 2022), 23 pages.

[7] DIREC. 2023. HERD: Human-AI collaboration: Engaging and controlling swarms of Robots and Drones - DIREC. https://direc.dk/da/herd-human-ai-collaboration-engaging-and-controlling-swarms-of-robots-and-drones/

[8] Mica R. Endsley and Debra G. Jones. 2016. Designing for Situation Awareness: An Approach to User-Centered Design, Second Edition. *Designing for Situation Awareness: An Approach to User-Centered Design, Second Edition* (1 2016), 1–373. https://doi.org/10.1201/b11371

[9] Yanwei Fu, Yanwen Guo, Yanshu Zhu, Feng Liu, Chuanming Song, and Zhi Hua Zhou. 2010. Multi-view video summarization. *IEEE Transactions on Multimedia* 12, 7 (11 2010), 717–729. https://doi.org/10.1109/TMM.2010.2052025

[10] Andreas Girgensohn, Frank Shipman, Anthony Dunnigan, Thea Turner, and Lynn Wilcox. 2006. Support for effective use of multiple video streams in security. *Proceedings of the ACM International Multimedia Conference and Exhibition* (2006), 19–26. https://doi.org/10.1145/1178782.1178787

[11] Maria-Theresa Oanh Hoang, Niels Van Berkel, Mikael B Skov, Timothy Merritt, and Timothy 2022 Merritt. 2022. Challenges Arising in a Multi-Drone System for Search and Rescue Challenges Arising in a Multi-Drone System. (2022). https://doi.org/10.1145/3546155.3546653

[12] Nina Hollender, Cristian Hofmann, Michael Deneke, and Bernhard Schmitz. 2010. Integrating cognitive load theory and concepts of human–computer interaction. *Computers in Human Behavior* 26, 6 (11 2010), 1278–1288. https://doi.org/10.1016/J.CHB.2010.05.031

[13] IGI Global. 2023. What is Visual Semiotics | IGI Global. https://www.igi-global.com/dictionary/visual-semiotics/48984

[14] Palash Yuvraj Ingle, Yujun Kim, and Young Gab Kim. 2022. DVS: A Drone Video Synopsis towards Storing and Analyzing Drone Surveillance Data in Smart Cities. *Systems 2022, Vol. 10, Page 170* 10, 5 (9 2022), 170. https://doi.org/10.3390/SYSTEMS10050170

[15] Richard E. Mayer and Roxana Moreno. 2010. Nine Ways to Reduce Cognitive Load in Multimedia Learning. *https://doi.org/10.1207/S15326985EP3801_6* 38, 1 (2010), 43–52. https://doi.org/10.1207/S15326985EP3801{_}6

[16] Shaohui Mei, Genliang Guan, Zhiyong Wang, Shuai Wan, Mingyi He, and David Dagan Feng. 2015. Video summarization via minimum sparse reconstruction. *Pattern Recognition* 48, 2 (2 2015), 522–533. https://doi.org/10.1016/J.PATCOG.2014.08.002

[17] Tao Mei, Bo Yang, Shi Qiang Yang, and Xian Sheng Hua. 2009. Video collage: Presenting a video sequence using a single image. *Visual Computer* 25, 1 (1 2009), 39–51. https://doi.org/10.1007/S00371-008-0282-4/METRICS

[18] NASA. 2020. TLX @ NASA Ames - Home. https://humansystems.arc.nasa.gov/groups/TLX/

[19] Zena O'connor. 2013. Colour, Contrast and Gestalt Theories of Perception: The Impact in Contemporary Visual Communications Design in Wiley Online Library. *Wiley Periodicals, Inc. Col Res Appl* 40 (2013), 85–92. https://doi.org/10.1002/col.21858

[20] Robotto. 2023. About | Robotto. https://www.robotto.ai/about

[21] Ben Shneiderman. 2022. *Human-Centered AI.* Oxford University Press. 0–377 pages.

[22] Tom Streeter. 2012. Definitions of Semiotic Terms. https://www.uvm.edu/~tstreete/semiotics_and_ads/terminology.html

[23] Zhida Sun, Mingfei Sun, Nan Cao, and Xiaojuan Ma. 2016. VideoForest: Interactive visual summarization of video streams based on danmu data. *SA 2016 - SIGGRAPH ASIA 2016 Symposium on Visualization* (11 2016). https://doi.org/10.1145/3002151.3002159

[24] Philip Tqulnlanu and Richard N Wilton. 1998. Grouping by proximity or similarity? Competition between the Gestalt principles in vision. *Perception* 27 (1998), 417–430.

[25] Chao-Ming Yang and Tzu-Fan Hsu. 2015. Applying Semiotic Theories to Graphic Design Education: An Empirical Study on Poster Design Teaching. *International Education Studies* 8, 12 (2015). https://doi.org/10.5539/ies.v8n12p117

[26] Saelyne Yang, Jisu Yim, Juho Kim, and Hijung Valentina Shin. 2022. CatchLive: Real-time Summarization of Live Streams with Stream Content and Interaction Data. *Conference on Human Factors in Computing Systems - Proceedings* (4 2022). https://doi.org/10.1145/3491102.3517461

[27] Kexin Yi, Chuang Gan, Yunzhu Li, Pushmeet Kohli Deepmind, Jiajun Wu, Antonio Torralba, and Joshua B Tenenbaum. 2019. CLEVRER: CoLlision Events for Video REpresentation and Reasoning. (10 2019). https://doi.org/10.48550/arxiv.1910.01442

[28] Zhiwang Zhang, Dong Xu, Wanli Ouyang, and Chuanqi Tan. 2020. Show, Tell and Summarize: Dense Video Captioning Using Visual Cue Aided Sentence Summarization. *IEEE Transactions on Circuits and Systems for Video Technology* 30, 9 (9 2020), 3130–3139. https://doi.org/10.1109/TCSVT.2019.2936526