Imperial College London

Department of Electrical and Electronic Engineering

Final Year Project Report 2016

Project Title:	**Assisting Search and Rescue through modified video presentation**
Student:	**Dourtchev N. C.**
CID:	**00735545**
Course:	**EIE4**
Project Supervisor:	**Professor Robert Spence and Dr Mark Witkowski**
Second Marker:	**Dr Thomas Clarke**
Extra Credit:	**Dr James Mardell**

Abstract

This project evaluates different presentation methods of the video feed from aerial photography in the context of Urban Search and Rescue (USAR). The aim is to determine which modification will have the most success in facilitating the finding of "targets" within the video by an observer. Two user experiment were carried out. Subjects were given the task to find targets - 'lost' school children with a specific description - in an urban environment created with the help of a computer generated simulation. The first variable that is evaluated is the video presentation: Standard Moving Presentation (SMP), Serial Visual Presentation (SVP) and Live Serial Visual Presentation (LSVP). Results show that in a scenario that features many distractors (e.g. pedestrians), SVP generally performs better than the alternatives. The number of distractions was reduced for the second experiment where LSVP comes up to par with SVP overall, but outperforms it for moving targets. The second variable is the orientation of the cameras attached to the flying device. We find that video from cameras that are angled more towards the horizontal plane give a better result as it shows a better point of view on the "lost" targets.

Summary of Project Achievements:

- Assessment of current methods for Visual Search
- Invention of two new video presentation methods for Search and Rescue
- Development of a simulator to allow the testing of the new presentation methods
- Postulation and testing of hypotheses regarding the investigated presentation methods
- Organisation and execution of two user experiments where the test subject takes on the role of target spotter in a Search and Rescue simulation
- Based on a statistical analysis of experimental results, conclusions regarding the relative performance of four presentation methods in Search and Rescue scenarios
- Formulation of recommendations for the Search and Rescue community
- Identification of useful further research

Conclusions

- LSVP is superior to SMP and SVP for moving targets in a low distraction environment
- SVP is superior to SMP and LSVP in a high distraction environment
- Overall superiority between SMP or LSVP is dependent on the task
- LSVPT is superior to LSVP

Table of Contents

1 Introduction

Technology and communications have come a long way in recent years. Unmanned Aerial Vehicles (UAVs), as well as video equipment has reached an affordable price that allows organisations with limited capital to make use of aerial video photography to fulfil their purposes (Goodrich, 2008). In the field of Search And Rescue (SAR), it is becoming increasingly popular to make use of this technology in order to improve the efficiency of search operations (RAF, 2011).

Typically, an individual is assigned the role of sensor operator or "spotter" amongst the other roles (e.g. pilot). This trained individual is given the task to inspect the video feed originating from the downward-facing camera of the flying device, and finding the missing person or object (Goodrich, 2008).

The typical air search success rate of untrained personnel, given the task of finding a target situated on the ground from an aerial video feed, can be as low as 30% (Croft, 2007). Further research into the success rate of such tasks has shown that a trained "spotter" can only provide an improvement of 10% (Stager, 1975).

Considering that extensive research has been conducted in training "spotters" to be better at their task, we try to propose a way to modify the video capture and its presentation to the "spotter", aiming to increase the success rate of a SAR operation.

We propose the modification of two distinct variables concerning video capture. The first variable is the Viewing Mode (VM):

1. Standard Moving Presentation (SMP) – typical live video, moving along the terrain, the "spotter" watches the feed as the UAV moves along the surface.
2. Serial Visual Presentation (SVP) – the "spotter" is presented with static images of the terrain that are taken as the UAV flies along. The image is presented to the observer for a specific amount of time until it is replaced with the next one.
3. Live Serial Visual Presentation (LSVP) – the "spotter" is presented with live video from a drone that is standing still above the surface, after a specific amount of time the drone and the frame move forward, and the next frame is displayed. The features within the frame are moving in real time, hence Live.

The second variable is the Number of cameras and their orientation:

1. Single camera with a moderately wide lens facing downwards

2. Two cameras facing outwards (providing an oblique view) – aiming to capture the profile (or height) of a person that is standing underneath the UAV, therefore providing additional information.

A computer simulation using the video game engine “Unity 3D” was used in order to create the scenarios and their features, and to create the video sequences on which the test subjects will be tested. This gave us great flexibility over how we wanted to model our scenario. The user experience study required the variables and separate experiments to be very well controlled and to introduce the least noise in the collected data. That is the main reason why we chose the computer simulation over the alternatives.

Two user experience studies were designed using the computer simulation that tests combinations of these variables in two different scenarios based on Urban Search And Rescue (USAR). Both scenarios feature the same environment, an urban river with its two banks, albeit different in their features. One is much more crowded than the other, creating significant distractions for the observer. The aim is to identify which presentation mode will yield the best operation success rate given the amount of distractions that are present in the scenario.

Thus, the general question that we want to answer is:

> **“What is the best way (amongst the ones we are testing) to present video footage for an USAR scenario in order to obtain the best success rate, given different levels of distraction present in the scene?”**

The results of the investigations aim to:

- Provide Search and Rescue services with an enhancement to the video streams collected by their aircraft, backed by experimental data, in order to achieve a better success at finding missing targets in a USAR scenario.
- Investigate how dependant the enhancements proposed are on the presence of distractions within the scene. And propose which enhancements should be used in what scenario.

Ultimately, our conclusions can contribute to further research as how video presentation can be improved in the context of SAR.

This project mainly builds on the work of Goodrich (2008) and Mardell (2013) that test the effects of SVP on target identification rate. After conducting tests on this previously proposed method, we further develop a new presentation mode (LSVP) influenced by the research of Holcombe (2010) and Abrams (2003).

After the Introduction, in Chapter 2, we discuss the existing methods for conducting Visual Search in the context of SAR, theory regarding the Visual Search and the way the Human Visual System (HVS) performs the task, and researched improvement in Visual Search on which our work will build on.

In Chapter 3 we talk about the Aims and Requirements that were set in the initial stages of the project.

Chapter 4 discusses the Analysis and Design of the different project components, namely the simulation platform used for testing and the different variables that were modified in order to test the proposed presentation methods.

Chapter 5 gives detail regarding the implementation of the SAR scenario simulator using the video game development software Unity 3D.

Chapter 6 and 7 go into detail regarding the user experiments that were performed to test the presentation methods using the testing platform. The procedures, hypotheses and results are stated and analysed. An evaluation to each experiment individually is provided.

Chapter 8 discusses the overall limitations and points of improvement of the project.

Chapter 9 gives the overall conclusion to the project. Including the project experience as well as a summary of the technical results.

Chapter 10 presents the reader and the department with further work that can be conducted on the subjects following this project.

2 Background

2.1 Existing methods

In the field of Search and Rescue, downward facing cameras are often used from flying aircrafts (e.g. helicopters) or UAVs to stream video to an observer, or “spotter” who is given the task to identify certain targets (Goodrich, 2008). These targets can either be missing or hiding humans, or even specific objects such as a missing aircraft. Most often, the video is streamed to the “spotter” in its original format, a moving / live picture across the terrain.

Figure 1.1: A drone with a camera installed (in red), belonging to Tinman RC, used for the search of a lost flying device in a crop field. (Tinman, 2014)

Figure 2.2: A helicopter with a video capture device (in red) belonging to the UK Emergency services, used for Search and Rescue operations. (UK Emergency, 2016)

Research into the success rate of such an operation has shown that an untrained observer would only find the targets 30% of the time (Croft, 2007).

There exists training for SAR observers that aims to specialise them in this task and improve the success rate of such an operation. One of the methods employed involves dividing the observed area into squares, with the observer fixating his vision on that particular square until he decides whether it is relevant or not. He then proceeds onto the next square (Stager, 1975). The objective of this method is to minimise the amount of time spent on moving your eyes across the area, and instead use it to fixate on specific regions.

A further investigation was conducted by Stager and Angus (1975) that looked at the eye movements performed by the subject whilst looking for a missing aircraft in a video footage. They concluded that there was no significant difference between the eye movements performed by a trained employee in comparison to an untrained one. However, trained employees had a 10% better success rate, justified by the factor of being accustomed to observing similar scenarios and knowing what to look for within the video capture. These conclusions suggest that the Human Visual System (HVS) cannot be trained to use an alternative method to improve its success rate for the given task. Given that extensive research has already been performed with the aim of improving an observer's search methods, our aim instead is to modify the video presentation to achieve a positive and improved outcome.

As previously mentioned, the most widely employed presentation for the video feed from an aircraft is the standard moving mode, where the observer watches the live video directly (RAF, 2011).

Figure 2.3: Shows an example of the widely used Standard Moving Presentation (SMP) of the video feed captured from a UAV

Research conducted by Khurana and Kowler (1987) investigates smooth pursuit – the act of following an object within a moving frame. The subjects in their experiment have to find a certain letter amongst others in a text that is moving along. Results show that when the text is moving along, the success rate of test subjects in finding the target letters deteriorates by 20%. This result can theoretically be transferred over to SAR video observation, indicating that a standard moving mode is not the most appropriate mode of presentation for SAR.

Burr and Ross (1982) found that the human HSV cannot find low-contrast objects moving at a faster speed than 32 degrees per second (o/s), a significant degradation in performance was observed at much lower speeds such as 2 o/s. Research shows the HVS adopting different strategies when observing still versus moving imagery. A sequence of short and stable fixations separated by quick saccades is associated with observing still images (Chan, 1996); (Tatler, 2003). Whereas, periods of smooth pursuit between which saccades are observed, is associated with moving imagery (Kowler, 2011). These different methods that the HVS employs for the specific cases makes it even more intriguing to investigate the results of an experiment comparing SMP and SVP.

2.2 Visual Search

Visual attention is a process employed by the Human Visual System (HVS) that selects an element out of all the visual information that is available to the human eye, and then brings it to the person's visual awareness (Wolfe, 2004); (Verghese, 2001).

Bottom up attention occurs when an object is brought up to the human's attention non-selectively, the HVS brings it into attention for its oddity or specific features that stimulate it. It is a process that takes the non-selective pathway in the brain because a person cannot consciously set himself up to look for it (Wolfe, 2007). It makes use of the parallel collection of processes within the HVS (Treisman, 1980).

On the other hand, top-down attention is strategic and requires an internal cause or origin. The HVS selects items of interest given the task that it was asked to perform by the brain. Typical criteria for selection would involve a combination of features (shape, colour, orientation, etc.), when the criteria is met, the object is brought to the person's attention (Carrasco, 1995); (Nakayama, 1986).

Visual Search

As Mardell (2013) suggests, these two types of attentions allow the target to be brought up to the HVS, but are insufficient to explain the identification process that has to occur during visual search. The combination of both attention processes will form a part of visual search. As Mardell (2013) remarks "A further series of complex processes is required in order for the HVS to attend to an unusual object or search automatically and methodically through a vast environment for a specific target."

Efficient (Parallel) and Inefficient Search (Serial)

Triesman and Gelade (1980) identify efficient and inefficient search in order to explain the previously mentioned series of processes. The process of efficient search employs bottom-up attention where the human's vision is scanned in parallel. The number of items or features in sight does not affect the speed at which the search is conducted (Humphreys, 1993); (Woodman, 2003). Inefficient Search however, is known as a serial process which can be invoked by the HVS. It uses a serial method to scan a visual stimulus and look for features that match the description that the human is looking for.

Classification of Urban Search and Rescue in the context of Visual Attention

It can be argued that SAR involves a third type of search process called Explicit Search as identified by Mardell (2013) "where the explicit, volitional control of eye movements is used to search a scene. Explicit search incorporates both efficient and inefficient search but attempts to control these processes through top-down attention." "For instance, explicit search is deployed by SAR spotters. They are advised to maximise their concentration out of the aircraft window by keeping eye movements short and fixations long." A more common example of explicit search might be identified in the act of crossing a street and looking out for cars.

For our task of USAR it can be argued that explicit search is used. The test subject employs Inefficient Search to scan the frame serially trying to find the specific characteristics that a target might feature, whilst getting his attention attracted efficiently through bottom-up attention when a target performs unusual or odd behaviour.

2.3 Researched Improvement in Visual Search

Further research exists into automated methods of human body detection and geolocation using colour and thermal imagery (Rudol & Doherty, 2008). Whilst very effective in a controlled and "expected" environment, these methods show a lower performance when unexpected or more complex conditions occur that cannot be compensated by the system. Typical examples are: unknown colour of the target, possibility of the target hiding or the heat map being obstructed by objects. Additionally, the presence of many similar targets (for example in an urban scene), can confuse the system. Such conditions can cancel out the benefits of this method and can result in an unsuccessful search.

Despite the challenges that are faced when looking for a target in aerial photography, "spotters" are still considered the best and most effective way to conduct a search (Croft, 2007), regardless of the limited time, resources and other difficulties faced with human work (Stager, 1975). The reason being is that machine learning has not yet developed enough to be able to swiftly adapt to SAR scenarios and all the unexpected factors that can come into play. For these reasons, we decided to conduct our research into improving the interface for "spotters", rather than working on automated systems for SAR.

Goodrich et al. (2008) researches different methods for visual presentation "involving temporally localised mosaics". In a Wilderness SAR scenario, test subjects are asked to identify red umbrellas with the different mosaic presentation methods. The findings show that subject were 43-45% more likely to identify the targets using the presentation methods proposed by Goodrich.

The PhD thesis of James Mardell (Mardell J. P., 2013) from the Department of Electrical and Electronic Engineering at Imperial, conducts similar research regarding the ways in which the presentation of the video feed can be modified. The scenario of Wilderness Search and Rescue (WiSAR) is used where the video is built from still satellite images. Four presentation methods were evaluated and compared to the standard moving video that is widely used. These are: Serial Visual Presentation (SVP), identical to SVP used in this report, Segmentation-based Inspection, Enlargement-based Inspection and Gaze-contingent displays. The results show a clear improvement in the test subjects' success rate of finding targets when using SVP. However, the other modifications show no statistically significant improvement. Specifically the paper, A Comparison of Image Inspection Modes for a Visual Search and Rescue Task (Mardell et al., 2012) describes the improvement in target location using the SVP mode. Additionally, they find that this improvement is supplemented by an increase in false positives, cases where the observer identifies an object that is not a target. We engage in testing the same SVP mode as J. Mardell, albeit in an Urban rather than Wilderness scenario, and with moving objects within the scenario (as opposed to still images), achieved with the help of the computer generated simulation.

This problem related to visual search is worthwhile to solve because previous research suggests that different video presentation modes have the potential to improve target identification rate. However, because the results are not conclusive enough, these methods are still not implemented by SAR organisations. The aim of this project is to further develop the ideas previously mentioned by Goodrich et al. (2008) and Mardell (2013) by firstly testing their proposed method and then developing new video presentation methods (LSVP, LSVPT) that will be verified and proposed to SAR services as an alternative the Standard Moving Presentation (SMP).

3 Aims and Requirements

In the initial phase of the project significant time was invested into specifying the aims and requirements for the project. The objective of testing different viewing modes and the experimental nature of the research required very clear guidelines as to what we wanted to achieve, this was coupled with the requirements of a platform on which to create the scenario and the USAR video, which served as a base for the user experience studies.

Viewing Modes

First and foremost, we aim to test different viewing modes for the presentation of the video feed. Ideally we liked to start off with a viewing mode that had already been studied. In that way we would make sure that the basis for our investigation relies on investigations with very good credibility. Consequently, by comparison and analysis, we would be aware whether our testing platform and other variables have been correctly designed. With our experience and findings, we would be able to further develop our model by stepping on the solid base already established.

Furthermore, we aim to develop and test a new viewing mode that has not yet been investigated. This is later referred to as the Live Serial Visual Presentation (LSVP).

An important criteria to consider when designing the viewing modes is to make sure that they are realistically implementable. The project aims for its finding to be practical and ultimately implementable by SAR services, considering their typically limited funds (UK Emergency, 2016) (Tinman, 2014).

Experiment Platform

The question of what testing platform to use was left very open at the beginning of this project. However, some recommended characteristics were set-up that later helped us decide to use the computer generated simulation amongst the other considered variants (described later).

The testing platform should be flexible, it should not require great effort to modify the scenery, the targets or any other features of the environment. We expected a lot of modifications to be made along the way, and many little details to be changed, such as target appearance and location, distractions, and terrain characteristics. A flexible and easily modifiable scenario will be a great advantage given the limited time available to complete the project. It would allow us to easily implement our ideas and make any quick modifications if needed because surprises can always come up and we want to be able to deal with them swiftly. Having an easily modifiable scene will also let us design our variables (viewing modes, distractions) very specifically and, more importantly, to make them as independent of each other as possible so that they do not introduce noise in the test results. For example, changing the viewing mode might create a tilt in the camera which might introduce more / fewer distractions than is the case in a different viewing mode.

It is recommended that the testing platform builds on previous work, rather than repeating it or going too far off. This would allow us to make better comparisons of our work with other research conducted in the area, and draw important conclusion.

This brings us to the realism in the scenario that can be created using the platform, in order for it to be close to existing research, it should present a similar level of realism and appearance. At the same time, it should be as realistic as possible. By making it realistic, we contribute to the credibility of our findings and can prove that they can be used in the real world for SAR operations.

Finally, the platform should enable and facilitate any further work or, in other words, it should be reusable. We expect many ideas for future works to come out of this project that may inspire other students' Final Year Projects or even PhDs.

User Experience Studies

The project relies on conducting user experience studies in order to identify be the best viewing mode. For this reason we set clear goals in that aspect. These often overlap with the requirement for the testing platform therefore we will only outline what has not been mentioned already.

The experiment should be user friendly. Keeping in mind that we were not sure who will be our test subjects in the beginning, we aimed at creating an experiment that tenders to all kinds of people: it should not require any kind of excessive physical movements, it should not risk to produce dizziness or sickness in the test subjects and it should not require the test subjects to have a characteristic that is uncommon (test subjects should not be required to be SAR trained or have perfect eye-sight – having glasses should be acceptable).

The experiment should be repeatable. It must be exactly reproducible between different test subjects. The scene that test subjects observe should be very well controlled and be identical between subjects performing the same task. The experiment should not be too long, or too short, so that it gives time for the person to get used to the task, but does not let them get bored and lose focus after a certain amount of time. It is recommended that the experiment requires the direct effort of only one individual so that additional variables are not introduced regarding the cooperation between individuals.

Results

The results that we take out of the user experience studies should have statistical significance in order to give them credibility in the scientific world. All data must be analysed as carefully as possible in order to draw the most conclusions as we can. Finally, the results must be provided in an adequate and understandable format and provide clear guidance on how they can be used in the real world to improve the existing SAR systems depending on the problem that they are facing.

4 Analysis and Design of Project Components

4.1 Choice of simulation platform

One of the initial tasks was choosing the platform to build the scenario on. This was generally the most important choice we had to make in the initial stages and significant thought was put into it, as everything else depended on it. We will hereby describe the three approaches that were discussed and partly implemented as well as their advantages and disadvantages and then explain why we finally chose to use a computer generated simulation using game engine software as the platform.

<u>Satellite Imagery</u>

In recent years there has been an increasing amount of detailed imagery from satellites and aircrafts of the earth's surface, especially urban areas. We would find images of a suitable scene (e.g. Figure 1) and edit them into video streams using specifically tailored software, similarly to what has been done by our colleague Dr Mardell (Mardell J. P., 2013).

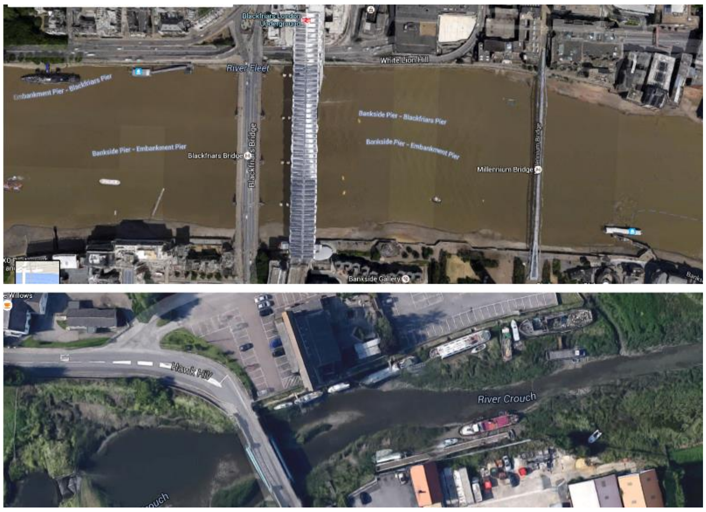

Figure 4.1: (TOP) Aerial View of the River Thames; (BOTTOM) Aerial View of the River Crouch; both taken from Google Maps as an example.

When generating the video feed, we will use different modifications to the presentation. Targets will be placed in the video and will have to be searched for by our test subjects to determine which viewing mode is best for target search. We would have the opportunity to reuse the methods employed by James Mardell and build on them to create our scenario and tests. We will save valuable time that can be spent on the experiment.

On the other hand, there are a number of disadvantages. It is hard to find high quality video footage with high detail of the terrain, this would limit the video to a minimum height from the ground, because as we go lower, the quality of the image will deteriorate. Creating moving targets and distractions will be a very time consuming process as it would likely involve modifying each single frame of the video manually. This signifies that the platform might not be flexible enough for our studies which contradicts one of the important points in our requirements, this will result in a lot of lost time in tailoring the scenario to our needs (target placement, distraction placement, etc.). Furthermore, the picture will most probably have low depth – completely 2D (because it is hard to find high quality images), this will result in a less realistic scenery, and consequently we will not be able to modify the camera angle.

Gathering Video Footage of the Scenario

This approach consists of organising a real scenario and gathering actual aerial footage in the form of video. It can be conducted in two ways:

1. Find a suitable location to fly our personal remotely controlled helicopter and get aerial footage of the desired terrain. Placing targets in suitable locations before filming, or adding them later during the post-processing of the video.
2. Organise to attend in a drill SAR operation during which we will be required to hire UAV operators to perform the filming.

As much video footage as possible will be recorded during the drill operation and will be edited to become our experimental footage. The advantage of this method is that the video footage will be from real physical UAVs, and an almost perfectly realistic (drill) search and rescue operation. This is as close to reality as possible given the circumstances. If conducted correctly, this experiment has the highest potential of credibility.

However, there are significant disadvantages to this approach. Firstly, many parties will be involved (UAV pilots, SAR personnel, etc.) which means setting up the experiment will be time consuming and unexpected circumstances may arise that can greatly decrease the validity of our experiment. We will have no direct control over how the video footage from the drones is captured because third parties will be controlling them. Moreover, we will only get one or two opportunities to gather the required video footage, if they do not go according to plan we might not have enough time to fall back to a secondary viable plan. Finally, we have no control over the location and environment of the experiment nor the placement of targets.

Game Engine Based Environment

The environment will be based on a Game Engine platform. Software will be used to assemble different assets (3D models) into a suitable environment for the test. A town will be created, with almost any features that we desire, as long as the cost-benefit is viable because not all assets are free. Cameras will be programmed to fly above the river in any of the viewing modes that we target to test. In real time, the subject will observe the video footage from the UAV and try to identify as many targets as possible within the simulated, "living" scene.

The main advantage of this platform is the flexibility which it provides. We are completely free to choose and design the environment, strategically place targets and distractions, and any additional features within the scene. Animated and moving characters and / or objects can be implemented to serve as distractions. Controllers can be created for test subjects to provide functionalities such as controlling the UAV. It is also be very easy to make improvements and iterations over the design whenever we desire or when we identify a flaw. Additionally, using this method we are independent from third parties (except test subjects for the experiment) which will make organisation easier.

The only significant disadvantage is that we will not be using a physically existing environment for the scenario which risks our credibility and can potentially expose us to criticism. However by using high quality graphics this effect can be reduced. Moreover, it must be noted that a lot of scientific credibility can be gained by correctly isolating the variables and external factors using the provided flexibility.

Final Choice and Reasoning

We decided to use the computer generated simulation as the platform for the development of this project, as the advantages were considered to far outweigh the disadvantages in comparison to the other approaches.

The main advantage that influenced the decision was the flexibility of this option and all the possibilities that it provides. Most importantly we want to have complete control over the variables that are being tested in order to introduce the least noise into the test data. With this platform, we would be able to set-up the terrain, cameras, targets, distractions and just about anything in the exact way we required. Additionally we will be able to work on it at any time from our own computers without depending on third parties and locations. The research would therefore have the potential to yield a wider range of conclusions.

Furthermore this represents a perfect platform for further work on the subject, meaning that future development can be built directly on top of the work completed in this project, rather than having to develop SAR scenarios from scratch using the other described approaches. The ideas will be discussed under Chapter 9 – Conclusions and Further Work.

4.2 Choice of Software

Several game engines were considered, tested and analysed before choosing the best one to construct the SAR scenario. The considered ones being Unreal Engine 4, Cry Engine 3 and Unity 3D.

Unity 3D was chosen as the Game Engine for the project. It provides satisfactory function for our environment, not as much as the other two options however, it does not feature a steep learning curve and work will be more efficient. It is compatible with all types of operation systems and supports all formats of assets. Additionally, it features great community support which will further facilitate our work. And finally it offers a free version where most of the needed functionality is supported.

4.3 Type of Scenario – Choice and Design

The initial specification of the report indicated that we should work on Urban SAR, mainly due to the fact that Dr Mardell (Mardell J. P., 2013) previously worked on Wilderness SAR for his PhD. We were therefore interested in researching urban scenes as well. Additionally, the supervisors were interested in looking at urban rivers and their banks, this is quite a popular SAR scenario because of the amount of traffic and danger that can be found. A large proportion of SAR operations are actually conducted along urban rivers such as the river Thames. Often people find themselves falling into the river, or walking along a river when they are lost. Additionally, floods create an increased risk in river regions and SAR operations have to be launched to rescue victims (Watson, 2013).

It was not completely out of the question to work on Wilderness SAR either. The main benefit of working with USAR is that it would give us a good comparison of which presentation modes work best in which type of SAR scenario. The two types of SAR greatly differ from each other in terms of the amount of traffic and distractions that can be observed. For example trees and houses represent a big limitation in the wilderness scene, whereas the main limitation in an urban environment comes from the large number of individuals present in the scene.

After completing Experiment 1, we found that the newly designed viewing did not perform as well as we expected. We therefore decided to modify the amount of distractions in the scenario for Experiment 2, following the theory that this might have been the cause. We strongly debated what kind of scenery to use for Experiment 2, where the distractions would have taken a low value naturally. Two propositions came up: designing a completely different Wilderness scenario using the game engine software, or reducing the amount of distractions in the already existing Urban SAR scenario. We chose the latter option for several reasons. Firstly, a Wilderness scenario would be vastly different from the urban one in terms of features, resulting in new and unknown factors affecting our observations and results. Ideally we wanted only the variables of “video presentation modification” and “amount of distractions” to affect the result rather than others that might not be easily identifiable. Secondly, creating a completely new scene would have cost much more time comparing with the time taken to simply reduce the amount of features present in the already existing scenario.

Realism was a main priority. It would be absolutely vital to touch up all the little details that make an urban scene exactly that. Human and vehicle traffic, streets, bridges, boats, bus stops, pedestrian crossings, street lights, postal boxes are just some of the examples of the features that we prioritise. Careful analysis of real city streets and rivers was conducted including the river Thames in London, the river Amstel in Amsterdam, and the river Chao Phraya in Bangkok. These were used as guidance on how to construct our scene. Unfortunately we were limited by the available 3D models (assets) that could be found for free or a reasonable price online, therefore our scene inspired itself from real urban scene as much as possible within this constraint.

4.4 Tested Variables

4.4.1 Video Presentation Mode

Standard Moving Presentation (SMP)

The first Video tested Presentation will represent the group in our experiment. The video is presented in its standard form – live, as the camera moves along the terrain (Figure 1.3). SMP is the most widely spread, if not the only presentation currently used in SAR. For this reasons we will compare all the modified modes to it.

Serial Visual Presentation (SVP)

The proposed method of SVP will investigate whether it is better to show a series of still images (as shown in Figure 4.2) captured from the video, as opposed to the moving video – SMP.

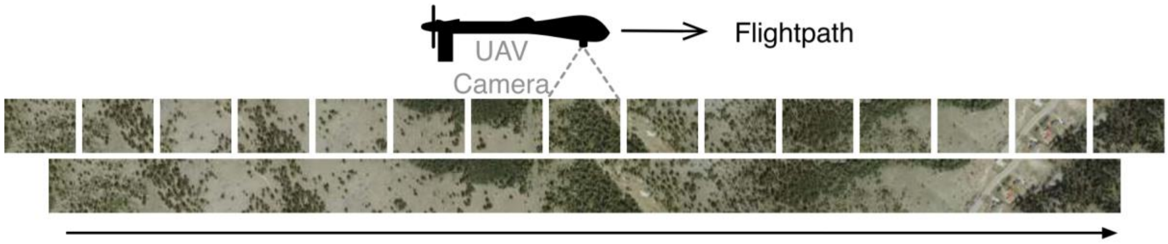

Figure 4.2: The terrain is split into numerous snapshots (top) that are taken in real-time as the drone moves along its flightpath. The snapshots are still pictures and no movement can be seen (copied and modified from (Mardell J. P., 2013)).

As mentioned in Chapter 4.2, the research of Burr and Ross (1982) found that the human HSV cannot find low-contrast objects moving at a quicker speed than 32 o/s, and that significant degradation in performance was observed at much lower speeds such as 2 o/s. In addition, their finding that the HVS employs different methods to conduct visual search when observing SMP vs SVP footage makes it even more intriguing to investigate the results of an experiment comparing these two presentation modes.

Additionally, Goodrich et al. (2008) and Mardell (2013) – Chapter 4 – investigate this method on visual search and conclude that it shows improvement for scenarios different than USAR. To confirm their findings, we will put this viewing mode through a series of tests using our USAR scenario created in a computer generated simulation.

SVP has an additional advantage that it uses less data and therefore transmission rates need not be high. This also means that the quality of the transmitted picture can be much higher.

Live Serial Visual Presentation (LSVP)

LSVP is similar to SVP in terms of showing a sequence of frames from the original video, however the images are not still but moving. The test subject observes an area of terrain in real time (with the world moving underneath), then after a certain amount of time, the view skips over a certain distance forwards, the process is repeated.

Even though the principle is simple in the visual theory, its practical implications must be explained. By observing Figure 4.3 below:

1. Green UAV remains static over a tile of terrain and records video.
2. When Green UAV finishes filming it will now fly forwards to its new assigned tile, skipping one tile of terrain. Meanwhile, Blue UAV films over its own tile of terrain.
3. When Blue UAV finished filming it will now fly forwards to its new assigned tile, skipping one tile of terrain. Meanwhile, Green UAV films over its own tile of terrain.
4. Steps 2 and 3 are repeated until the whole terrain is covered.

The main disadvantage of this method is clear. Two UAV's / drones will be required that will need to stop and go periodically.

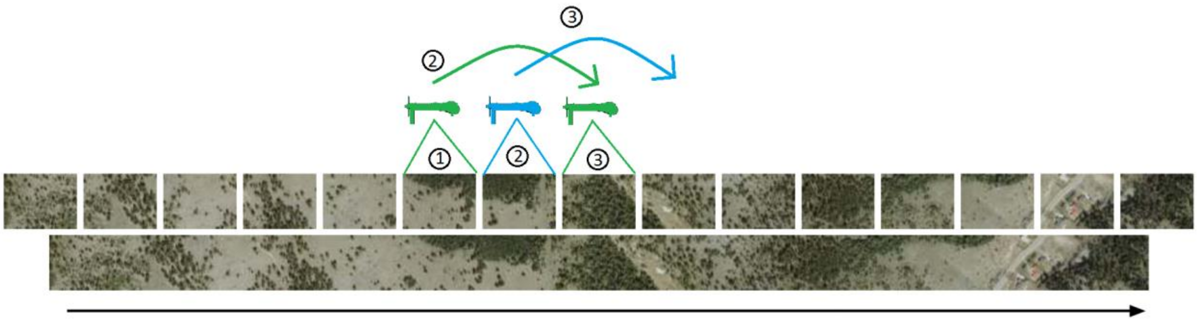

Figure 4.3: The terrain is split into numerous frames (top) that contain live video. These are shown one after the other in the sequence. The practical implications of this are explained in the text above (adapted from (Mardell J. P., 2013)).

However, a potential advantage is gained over the SVP method, movement within the currently displayed tile is not discarded. The effects of motion on visual perception have been widely investigated in the field. Holcombe's (2010) research investigates whether an unexpected change in direction of an object attracts attention. This is explored through a tracking task where grating patches change smoothly and semi randomly in position and orientation. Test subjects attempt to track the orientation of two of them. The stimuli disappear and one of the two gratings is queried for its orientation. The result shows that simple motion does not affect the correct answer given by the subjects, however, when the target collided with an invisible wall and changed direction the test subjects identified its orientation better. The experiment was repeated with a visible wall and now the change in direction had no effect. We can relate this unexpected motion to unexpected or odd human movement in a Visual Search context and verify whether unexpected / odd target movement will result in a higher identification rate.

Moreover, Abrams (2003) conducted an experiment to test whether motion and its onset attracts attention. Test subjects are asked to identify target letters in displays that contain targets and distractors. "*The results show no advantage for moving letter among static ones, but there is an advantage for objects that had recently started moving despite the fact that motion was uninformative in the context.*" Additionally, "*detection of target letters was found to be independent of the number of distractors in the display if the target had undergone motion onset*".

The fact that the conclusion that simple motion does not attract attention was drawn out of the context of SAR, calls for us to verify whether the same result applies to our investigation. We will investigate this by comparing the results for SVP and LSVP for moving SAR targets.

Live Serial Visual Presentation with Rotation (LSVPR)

LSVPR is again similar to LSVP, however it does not require two UAV's to be performed. The idea is that as the UAV flies forwards the drone rotates its camera backwards. Therefore a static object will remain at the absolutely same position throughout the frame but the angle of view will change, which causes a strange visual phenomenon.

The main advantage of LSVPR is that only one UAV will be required, whilst the disadvantage is that the view that it provides is not very intuitive and has the tendency to cause dizziness, an effect that we wanted to avoid in a public experiment. After implementing this presentation mode and analysing the implications, we decided to only use LSVP instead of LSVPR. The final three viewing modes used were therefore SMP, SVP and LSVP.

4.4.2 Display Modes

The second proposed variable was display modes. Or in other words, what kind of camera feeds to show on a single display monitor used by the spotter. This idea was a consequence of the initial planning to also include a boat that will float along the river, this would represent the SAR boat that performs its function within the river boundaries.

Figure 4.4 shows two of the display modes considered.

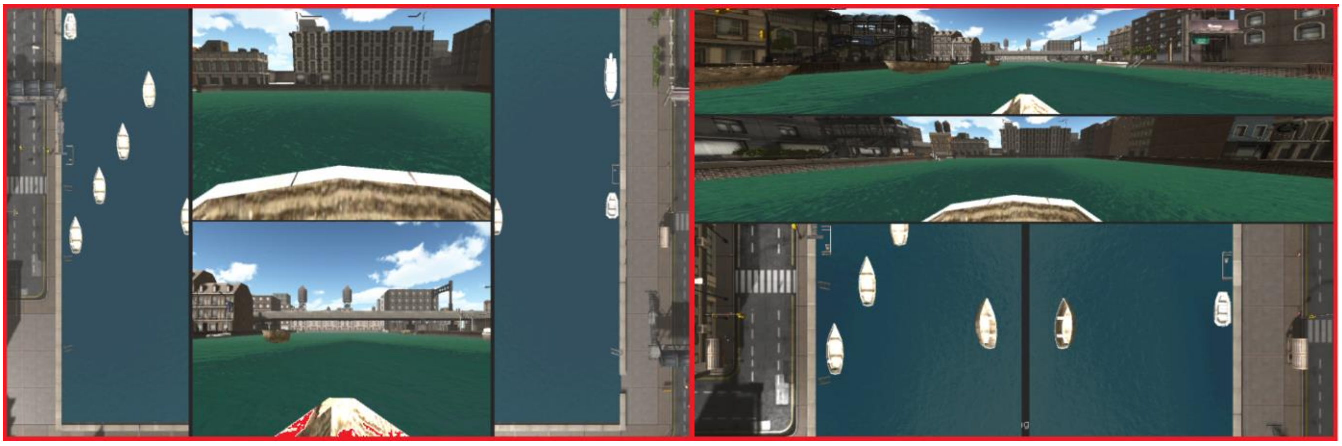

Figure 4.4: Shows two of the display modes that were considered and implemented for the experiment. The concept is to present both the view from the UAV as well as the view from the boats – forwards and backwards (as seen in the diagrams).

It was suggested that these additional viewpoints would give extra information and give better chance to spot the targets. On the other hand, involving this variable presented us with several complications. Most importantly, distinct viewpoint showed different areas of the terrain. For example from one viewpoint you might have a better view under a bridge, whereas from the other viewpoint you might get a better view behind boats parked on the bank. This means that there will be complications in the implementation of targets that are equally visible from all viewpoint, in order to avoid bias due to target design.

Additionally, a combination of viewpoint on the display meant that a test subject's attention would be significantly divided amongst the different displays. Some subjects might find it more intuitive to look at one display than the other, making them automatically more likely to spot targets that are exclusively seen from this viewpoint, and vice versa for another viewpoint. This factor would be augmented if we started to combine different Viewing Modes with different Display Modes.

Generally, we concluded that including different viewpoints would introduce variables that we cannot easily control and will cause further problems down the road. It will be inefficient to go down that route of research given the limited time available.

4.4.3 Level of Distractions

The results of Experiment 1 indicated that, relative to SL and SVP, LSVP does not facilitate the spotting moving targets given our Urban SAR scenario where we included a lot of distractions by design. We sought an explanation given the high expectations for this Viewing Mode, especially for moving targets since they should naturally attract more attention when they perform motion (in LSVP as opposed to SVP).

The first theory suggested was that moving targets distracted more than static ones. In LSVP the distractions were moving, whereas in SVP they were frozen in place hence, they did not distract the test subjects as much. This was verified with the data from Experiment 1 and hypothesis $H_{1.4}$ and was subsequently rejected.

The second theory suggested that the presence of many distractions (both moving and static), diminished the effect of bottom-up attention that should have been provoked by the targets' movement. Or in other words, test subjects were too busy closely examining the distractions, consequently their attention was not attracted by the moving targets. If this were true, and we were to present test subjects with a new scenario with less distractions, this effect will be diminished and target movement will have a higher probability to attract the test subjects' attention and result in a successful target identification. Therefore, LSVP would have the potential to perform better than SVP if there were fewer distractions present in the scenario.

We therefore decided to conduct Experiment 2, where we would test the same Viewing Modes, but the scenario will feature fewer distractors.

4.4.4 Number Scenarios / Sequences

It was considered whether we should have each test subject go through each of the 3 viewing modes in sequence on different scenarios, as opposed to being tested only on one scenario with one of the viewing modes. The former would require 3 different scenarios to be created, with different target locations.

The advantage of that method would be that we would be able to compare how each subject performed given the different viewing modes. We would apply a method such as ANalysis Of VAriance (ANOVA) for testing the data from 3 different groups (for each scenario) which would treat the variable of different scenario as independent. Overall this will result in a more efficient experiment, obtaining a larger sample base.

However we considered that there are significant disadvantages for our application. Firstly, we would have to hold up test subjects for an extended period of time in order to get them through all 3 scenarios. Given that we expected to conduct the test at the Imperial College Festival (public event), where people were usually in a rush to see all expositions, this was not suitable. Additionally we suspected that there will be slightly less motivated test subjects in the experiment who would lose focus at some point in the longer sequence that they will be presented with. Finally we decided to show only one scenario with one viewing mode for 3 minutes for each test participant.

4.5 Target Design

Designing the targets was closely related to designing the experiment itself. They had to be compatible with the instructions that would be given to the test subjects before they conducted the user experience survey. Instructions had to be clear and specific enough to ensure that every test subject knew exactly what to do and, more importantly, different individuals must obtain a similar understanding of the task. We chose "lost children wearing a red top and blue trousers" as the targets. Meanwhile, human distractions would typically wear a red top but different coloured trousers, or vice versa.

Additionally, each target having a specific, meaningful characteristic of their own enabled us later to draw more conclusions from the data. Each target was to be designed to be different from the others, in one way or another.

Static Targets

1. Sitting in the sidewalk areas
2. Standing in the sidewalk areas
3. Sitting in a moving boat on the river

Moving Targets

4. Running and hiding (within the sidewalk area)
5. Sitting in a moving object (i.e. boat) within the river area
6. Standing and waving within the sidewalk area

A drowning / swimming target was not included because the target's clothing would be submerged and we wanted the instructions to be clear in terms of the target specification.

Using these characteristics, we would be able to draw conclusions on how each variable introduced into the user experiment affects the finding of these specific classes of targets. In the Implementation section of the report we will include more information regarding target appearance.

It is also important to mention that targets were the main modification that took place throughout the two pilot studies that were conducted. The aim was that overall the finding of targets should occur in 50% of the cases, this would allow us to get a wider standard deviation of results and therefore be able to draw better conclusions from the data. In order to regulate the difficulty we modified the position and the colour of the target's clothes. For example if we wanted to make the target more visible (i.e. easier to find), we would position it to be more visible from the aerial view, make its clothes brighter and / or make it slightly larger. And vice versa if we wanted to make its identification more challenging.

The final design of the scenarios is discussed under Chapter 5: Implementation of the Environment. The final design of the experiments will be discussed under Chapters 6: Experiment #1 and Chapter 7: Experiment #2.

5 Implementation of the Environment

5.1 The Basics

When we were satisfied with our plan to create the Urban SAR scenario using the game engine software "Unity 3D" we had to start from implementing the basics and learning to work with the IDE (Integrated Development Environment).

In order to gain experience with Unity 3D, we started by watching tutorials on YouTube on how to make basic features such as Terrain, Water and Character Controllers. We proceeded to make a lake in order to learn how to create water and, later on, a river. Even though the free version introduces some limitations to the different features that can be added to water, it will still be enough to be functional for our design.

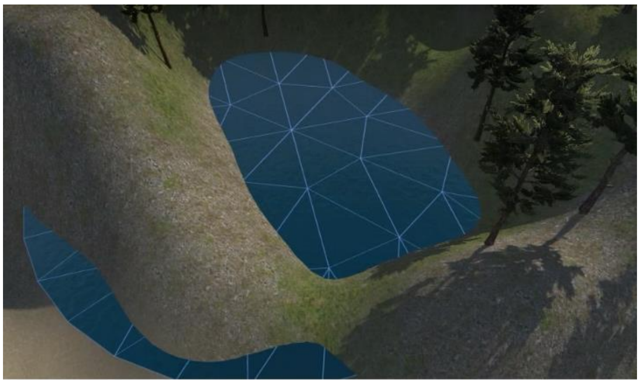

Figure 5.1: Working on one of our first projects to create water in a mountain environment

We then proceeded to learn how to make the water physically realistic in terms of object floatation, ripples, reflection, etc.

Urban Environment

In order to create the basic Urban Environment, we had to find assets that will help us build houses and urban features on our Unity project. We found a $10 package "Town Constructor Pack" that included most features that we would need to make buildings, streets, and various other urban features. As can be seen on Figure 5.2, buildings were made from building assets and positioned in a typical way for a modern city. Streets were placed in between the building. A demo river was created in the middle of town. In order to make the river banks realistic, numerous boats were positioned along them, together with ramps and small stairs for access to the boats. These are features whose purpose would be to obstruct the view of the observer in the boat.

Figure 5.2: Demo town environment created using the Town Constructor Pack

Additionally, we included small details such as bus stops, fire extinguishers, rubbish containers, benches, pedestrian crossing, traffic lights, trees, bushes and other decorations that would make the scene much closer to what would be seen in a modern western city.

5.2 Dynamic Elements

Additionally to all the basic details that were previously added, we required moving objects that will further add to the realism of the scene.

Pedestrians

Pedestrians are a main aspect of a realistic urban scene. To implement them we used the “Adventure Character Set” from the Unity 3D Asset Shop, even though these were not completely modern clothed models, they included meshes and joint models that allowed us to animate them easier in the next stage. This included the 10 characters shown in Figure 5.3. We also recoloured their clothes in order to introduce more variation amongst the pedestrians that will walk around the scene.

Figure 5.3: The 10 characters that were modified and animated to be used as pedestrians.

After animating the characters we had to create their movement paths along the river. We redesigned all the sidewalks to be straight all the way from the start of the terrain to its end, therefore we could write C# scripts for the pedestrians to walk along a straight line from beginning to finish. Four lanes for pedestrians were created, two going forwards, and two going backwards as seen in Figure 5.4.

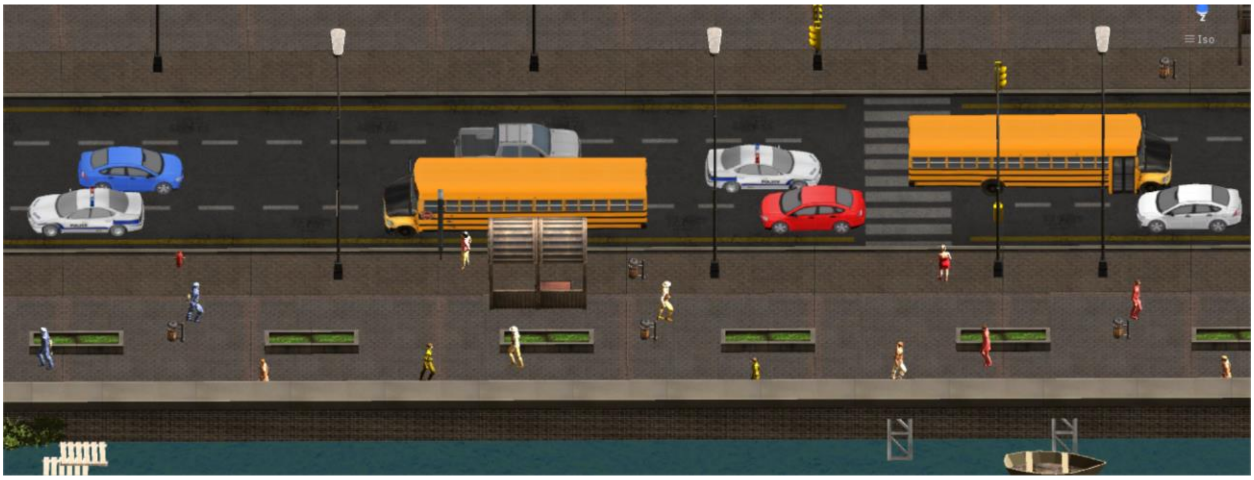

Figure 5.4: Shows the walking pedestrians and the cars features in the scenario

We decided to widen the sidewalks in order to give more space to simulate the pedestrians. The street for vehicles was put next to it. With the trade-off of limited UAV height (so that the people in the scene were visible), we considered that it is better to give priority to displaying pedestrians and moving cars, rather than static buildings, therefore buildings were excluded from the scenario. This also resulted in a more realistic scenario overall.

Static Humans

Apart from the walking pedestrians, we included static humans around the scene. Either standing or sitting (Figure 5.5), these would occupy boats, benches, but stops, and generally would be spread across the scene. 16 models were used, recolouring the clothing of some of the models resulted in a total of 40 different variations of static humans around the scene.

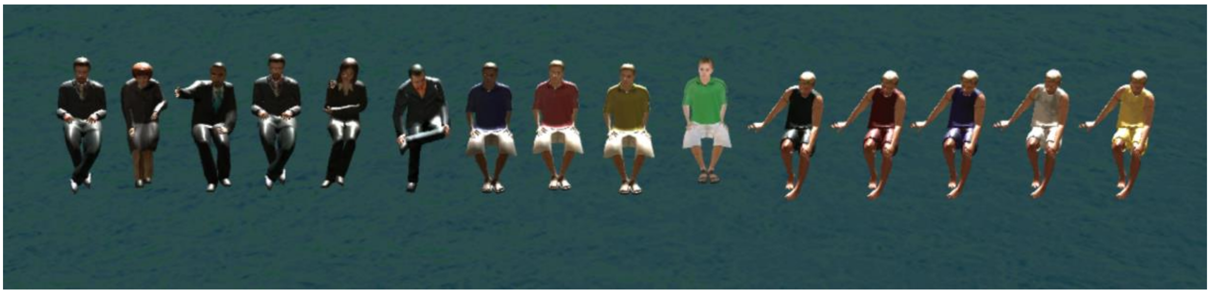

Figure 5.5: The 15 variations of sitting characters used in the scenario. Some of them are placed on boats, other on benches on the sidewalks on the river banks.

Boats and Cars

Similarly to how we implemented the pedestrians, we programmed moving boats along the river (Figure 5.6). Five different types of boats, each with four combinations of passengers sitting in them, resulting in a total of twenty variations of boat / passengers combinations in the river. Five movement channels were made along the river, three of them moving backwards, and two moving forwards. Again the movement was programmed using C# scripts.

Figure 5.6: The 20 variations of boats and sitting characters that would circulate along the river.

Seven different models of cars were programmed to move along the streets, one lane going forward and one going in the other direction. These are shown in Figure 5.4 above.

Distractions

Distractions take many forms in the scene and it could not be predicted how each one will affect the test subject before conducting the experiment. In general, the biggest distractions were considered to be humans or objects with red as their major colour, because it is similar to the colour in which the targets are dressed. Moving objects (i.e. boats, cars, pedestrians) also create a certain amount of distraction, which is why we put a significant amount of effort into designing them.

For Experiment 2 we removed a significant amount of the distractions in order to verify the effect on the viewing modes. Figure 5.7 shows how the scenario looked for Experiment 1 vs. Experiment 2.

Figure 5.7: Shows a comparison between the scenarios used for Experiment 1 - high level of distractions (top) and the one used for Experiment 2 - low level of distractions (bottom)

5.3 Targets

As previously mentioned in Chapter 4.5, we will use six targets for our scenario, where each target is designed in a specific way to provide valuable information regarding the variables that we test. More specifically the targets are designed as follows:

1. Sitting on a bench on the sidewalk
2. Standing under pedestrian bridge crossing on the sidewalk
3. Sitting in a moving boat on the river
4. Running from under a bridge to behind a bus stop
5. Sitting on a terrace that is attached to the bank and over the river
6. Standing next to a bus stop and waving

Target numbers 3, 4 and 6 are moving targets. All targets are shown in Figure 5.8 below.

Figure 5.8: Shows the 6 targets used for the user experiment, ordered from 1 to 6.

5.4 Cameras and View Modes

<u>Standard Moving</u>

For this mode, a camera was placed with a specific position and orientation at the beginning of the scenario. The camera was then programmed to move forwards with a velocity of 6 m/s along the x-axis, the axis along the river's length.

Serial Visual Presentation (LSVP)

The same camera as for the Standard Moving mode is used. It is now programmed to take snapshots every 6 seconds. These snapshots are saved in a folder and are then combined into a video using video editing software.

Live Serial Visual Presentation (LSVP)

For LSVP we start with a camera with a set position and rotation and we write a script with a counter that moves the camera forwards by one frame whenever 6 seconds have elapsed. This is a practical simplifications and does not reflect an implementation in the real world with UAV's for this viewing mode.

6 Experiment #1: High Distraction

We had the opportunity to participate at the Imperial College Festival on the 7th and 8th of May 2016 where we conducted Experiment 1. We had set up the apparatus behind a screened off area at the back of our assigned stand in order to avoid the test subjects from being distracted whilst taking the experiment. The apparatus consisted of a laptop that was connected to a large 4:3 screen used for displaying the scenario. A table was placed at the front of the stand where Professor Spence, Dr Witkowski and Dr Mardell attracted visitors and gave them more information about our work. Volunteers to take the survey were sent around the back where we conducted the experiment.

Figure 6.1: The front of the stand where Dr Witkowski presents our work to an interested individual. The area circled in red is the screened off section behind which our experiment was conducted.

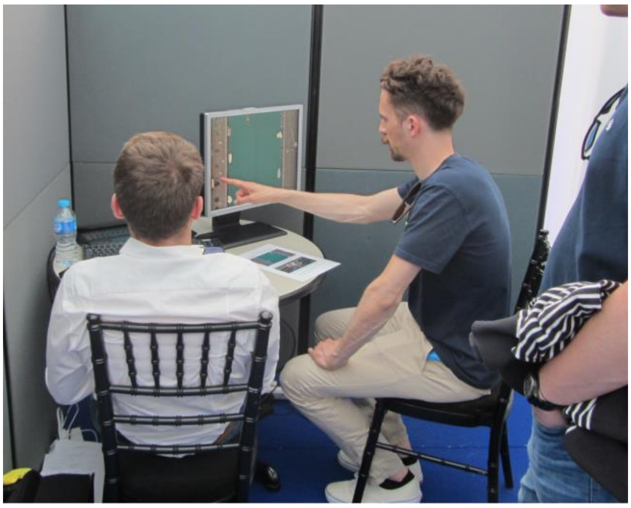

Figure 6.2: Behind the screened off section shown in Figure 6.1, a test subject takes the experiment, pointing to a potential target on the display monitor.

Due to the time constrains, the pilot study that took place prior to the festival was very limited. Five people were tested on the scenario and modifications to the targets were made as described in Chapter 4.5.

Experiment 1 uses the first scenario we created, with a high level of distractions.

We had prepared four variations of the scenario for this experiment, all with high level of distraction:

1. Standard Moving Presentation (SMP)
2. Serial Visual Presentation (SVP)
3. Live Serial Visual Presentation (LSVP)
4. LSVP with 2 cameras Tilted 20 degrees sideways relative to the vertical plane (LSVPT)

6.1 Experimental Procedure

As test subjects came along, we tried to balance the amount of test subjects for each presentation variant (SMP, SVP, LSVP and LSVPT). That way we would obtain a similar number of test subjects for each one.

Each test subject was given only 1 of the 4 scenarios that were available. That is because in all 4 scenarios the scene was identical, only the presentation modes were changed. If we were to show a test subject more than one of our scenarios, they would be better prepared on the second one.

The test subject was given a Questionnaire where they answer the first four questions regarding their Gender, Age, Eyesight and association with the Imperial College (Appendix A).

A sheet with instructions is given to the volunteering test subject, it briefly states the following (full version with pictures in Appendix B): A drone with a video camera flies over a river. You will be observing the video from its camera. Your task is to find 6 lost school children along the river. All the children are wearing a red top and blue trousers, see the example (shown in the picture in the full version). Whenever you think you see one of the children, point to that child on the screen. Don't forget to look along the river, as well as the sidewalks. The sequence will last about 3 minutes, please watch it to the end. Say START when you are ready.

Eight targets are placed in the scene in total, two of which are example targets in the beginning of the scene. The examples are clearly pointed out to the test subjects so that they are clearly aware of what they should be looking for. After the camera flies over the first bridge, the test subjects are not assisted.

Every time the test subject points on the screen, the experiment supervisor (Nikola Dourtchev) records on a piece of paper (Appendix C) whether a target was correctly identified (True Positive) or a non-target object was identified – False Identification (False Positive). Information regarding the position and identity of the falsely identified object is also recorded for further analysis.

After the experiment is over, the test subject is asked to answer the rest of the questions featuring on the Questionnaire under Appendix A.

Apart from the pilot study that was conducted prior to Experiment 1, it was required to treat the first 15 test subjects of Experiment 1 as part of a second pilot study. This resulted from the analysis of the first 15 sets of results we obtained, we concluded that the success rate of finding targets was either too low, or too high, depending on the specific target. We tried to normalise this using the method explained under Chapter 4.5 in order to achieve a success rate closer to 50%.

Apart from the 15 subjects used as a second pilot study, we obtained a total of 56 results during the two days during which the Festival took place. That is an average of one person every 10.99 minutes, including the breaks that the experiment supervisor took.

Additionally, we had a large range of age groups for the test subjects, from children under the age of 10, to older people over the age of 70. As expected, different age groups had different performance on the task. Therefore we tried to balance the age factor across the 4 scenarios that we were testing, meaning that we would aim to test a similar amount of people in each age group (specified in the questionnaire) for each scenario. During the first day of the festival we had obtained a larger number of older people due to the Alumni event that took place in the morning of that day. Subsequently we obtained more people in the older age groups for SMP and LSVP. We tried to compensate this by testing more young individuals (under the age of ~12) for SVP and LSVPT during the second day, this is justified by the fact that very young and older people obtained similar results.

Social Aspects

Participating at The Festival was a great experience, with many social outcomes in addition to the scientific ones.

We had the opportunity to present our department in a great light in front of the international audience. A wide range of volunteers were interested in our work, as well as the experiments that were set up at the stand. We put maximum effort to not only use the volunteers as test subjects, but also give them a better idea of what our research consists of. As a result, we are confident to have made an excellent impression on public.

There were a lot of learning outcomes for us in terms of communication skills with the wide variety of public that we were facing. We were required to adapt our behaviour to the age and maturity of the audience regardless of whether it was a young child or an Imperial Alumni. We had to combine fun and educational activities to satisfy both children, parents and college associates.

6.2 Hypotheses and Results

The results are summarised in the three charts below (Figure 6.3, Figure 6.4 and Figure 6.5). The hypotheses will be analysed one by one by observing parts of the data from each figure. Additionally we will discuss other interesting findings that can be observed from the collected data. After excluding outliers and subjects that were not fit to be considered for the statistical analysis, we obtained 14 test subjects for SMP, 16 for SVP and 22 for LSVP.

***$H_{1.1}$*:** Overall*, SVP will obtain a higher target identification success rate than SMP, using a high level of distractions (provided by the scenario in Experiment 1).

$H_{1.1}$ is aimed at confirming the findings of Mardell (2013) and Goodrich (2008) regarding SVP, summarised Chapter 2.3.

To test hypothesis $H_{1.1}$ we compare the overall results of target identification for SMP and SVP shown in Figure 6.3 below. We verify whether test subjects achieved a significantly better success rate at identifying targets using SVP in comparison to SMP. With SMP, test subjects correctly identified 47.6% (SD = 18.3%, N = 84) of all targets in the scenario. With SVP, subject correctly identified 62.5% (SD = 22.4%, N = 96) of all targets. These results are shown in Figure 6.3.

Statistical significance analysis using the t-test indicates that the hypothesis that $H_{1.1}$ is significant at a 5% confidence interval ($p=0.003<0.05$). We can conclude that a significant improvement in the successful identification of targets using SVP as opposed to SMP.

$H_{1.1}$ is confirmed. SVP performs better than SMP in the scenario with a high level of distractions.

Goodrich (2008) suggests that this is due to the static frames maintaing spacial context between each other, facilitating target detection. We further suggests that when each frame holds the screen for a specific amount of time, the test subjects can adopt a technique (speed and pattern) through which they search the static picture in order to have searched all of it by the time it changes to the next frame. In contrast, with SMP, the test subjects cannot be sure how fast they have to scan the picture in order to cover all areas. For example one starts from the top left corner, going down, across to the right, and up to the top right corner, then starts again from the top left. In SVP, when he starts again from the top left he is aware whether he is still in the same frame and had more time to search, or the frame switched before he could finish and he should speed up his search. Whereas in SMP, he cannot be exactly sure whether he has already searched the area in the top left corner and therefore cannot make a judgment on how to modify his search speed. In this context, LSVP should also benefit from the same advantage.

Overall* - Indicates that we take an average of the targets identified by each subject, and average that again to obtain the overall target identification rate for the specific viewing mode.

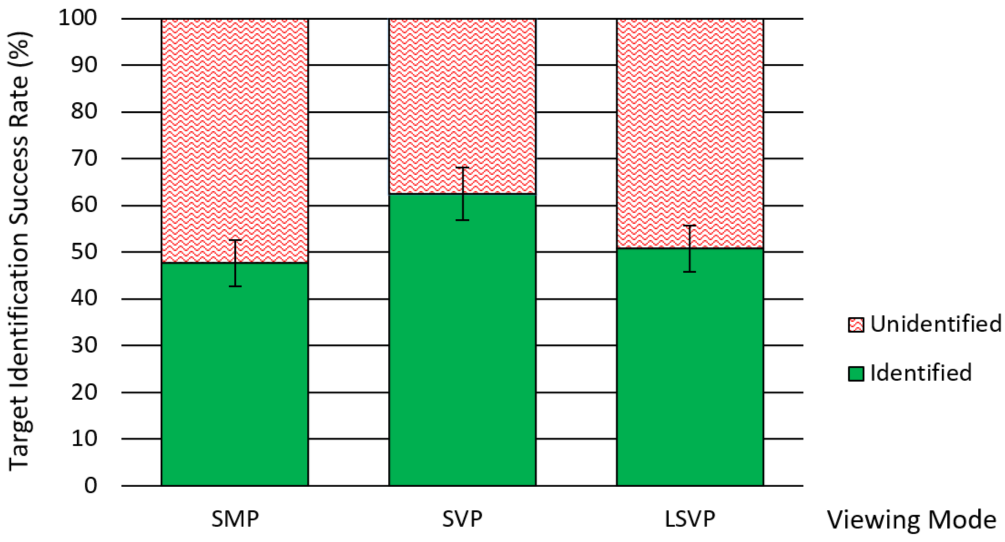

Figure 6.3: Summary of the overall target identification success rate (in %) for the 3 Viewing Modes tested in Experiment 1. The green bar represents the percentage of correctly identified targets, whereas the red bar represents the percentage of targets that were not identified. The black bar represents the standard error (at ±1.960).

$H_{1.2}$: Overall, LSVP will obtain a higher target identification success rate than SVP and SMP, using a high level of distractions.

$H_{1.2}$ is aims at to show that overall, LSVP should perform better than SVP and SMP because it combines advantages of both methods. Spatial context is maintained through a static frame (observed in SVP but not SMP), and live movement within the frame is maintained (observed in SMP but not SVP).

In order to test $H_{1.2}$ we compare the success rate for target identification rate with LSVP versus SMP and SVP in Figure 6.3 above. With SMP, test subjects correctly identified 47.6% (SD = 18.3%, N = 84) of all targets in the scenario. With SVP, subject correctly identified 62.5% (SD = 22.4%, N = 96) of all targets. With LSVP, test subjects correctly identified 55.3% (SD = 22.05%, N = 132) of all targets in the scenario. Statistical significance analysis using the t-test indicates that the hypothesis that LSVP is more successful than SMP is significant at a 5% confidence interval ($p=0.027<0.05$). Statistical significance analysis using the t-test indicates that the hypothesis that LSVP is more successful than SVP is not significant at a 5% confidence interval ($p=0.99>0.05$).

$H_{1.2}$ is rejected. The average of all targets taken into account, LSVP shows a significant improvement over SMP, but not over SVP.

Additionally we verify whether SVP is significantly superior to LSVP in the scenario. With SVP, subject correctly identified 62.5% (SD = 22.4%, N = 96) of all targets. With LSVP, test subjects correctly identified 55.3% (SD = 22.05%, N = 132) of all targets in the scenario. Statistical significance analysis using the t-test indicates that the hypothesis that SVP is more successful than LSVP is significant at a 5% confidence interval ($p=0.019<0.05$).

After comparing the overall performance of all 3 viewing modes, the only conclusion that can be stated is that SVP is superior to SMP.

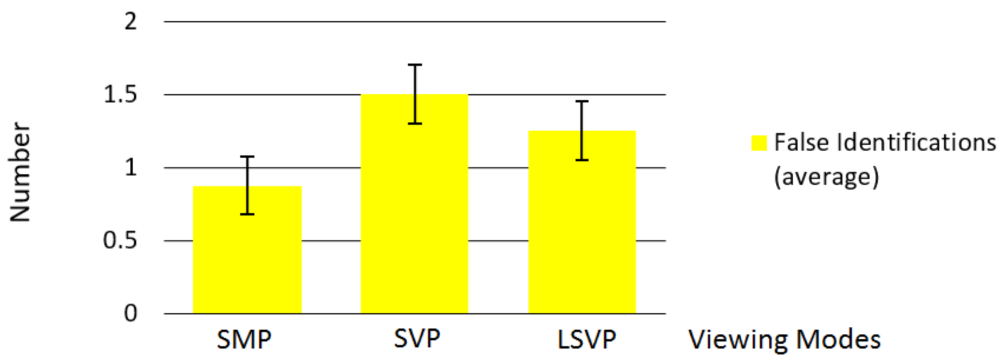

Figure 6.4: Average number of false identifications (False Positives) for the 3 viewing modes. The black bar represents the standard error (at ±1.960).

Additionally, we analyse the number of false identifications of targets. On average, 76% more false identifications occurred during SVP in comparison to SMP. Statistical significance analysis using the t-test indicates that the hypothesis that SVP experiences more false identifications in comparison to SMP is significant at a 5% confidence interval ($p=0.006<0.05$). We observe that the more correct identifications there are, in general there are also more incorrect identifications. The reason for this is suggested to be that in general the more the test subject points to targets, the higher the chance will be to find one, but there will also be an increased chance of it not being a target.

***$H_{1.3}$*:** LSVP will perform better than SVP and SMP for targets categorised as "Moving", using a high level of distractions.

$H_{1.3}$ is a hypothesis aimed at showing that movement within the USAR scenario contains valuable information that should not be discarded by simply using SVP. Specifically, LSVP should perform better for the targets characterised as "moving" in the scenario. In theory, target movement should attract the attention of the test subject and thus, increase the chance of target identification.

For $H_{1.3}$ it is important to remember that in the scenario itself 3 of the targets are moving. When viewed with LSVP, the same 3 targets are still moving, however with SVP, they are now frozen in place because SVP projects still images. We compare the results between LSVP and SMP for the moving targets 3, 4 and 6 in Figure 6.5.

With SMP, test subjects correctly identified 48.8% (SD = 22.91%, N = 42) of moving targets in the scenario. With SVP, test subjects correctly identified 53.3% (SD = 21.32%, N = 48) of moving targets in the scenario. With LSVP, test subjects correctly identified 51.1% (SD = 22.32%, N = 66) of moving targets in the scenario. Statistical significance analysis using the t-test indicates that the hypothesis that LSVP is more successful than SMP for moving targets is not significant at a 5% confidence interval ($p=0.74>0.05$). Statistical significance analysis using the t-test indicates that the hypothesis that LSVP is more successful than SVP for moving targets is not significant at a 5% confidence interval ($p=0.99>0.05$). Overall there exists no statistical significance that LSVP performs better than SMP or SVP for the targets categorised as "Moving". **$H_{1.2}$ is rejected**.

This indicates that in this scenario, a moving target (LSVP) is not more likely to be spotted by the observer than a static target (SVP) given the scenario with a high level of distractors. In Chapter 6.5 – Evaluation we consider the factors that may have affected this result.

***$H_{1.4}$*:** SVP will perform better than LSVP for targets categorised as "Static", using a high level of distractions.

$H_{1.4}$ attempts to verify the theory proposed in Chapter 4.4.3 where we suggest that moving distractions distract more that static distractions. The only changing variable between the two tested cases is that in SVP the distractions will be static, whereas in LSVP the same distractors will be moving. There will be no difference to the targets because they are "Static". Therefore $H_{1.4}$ compares the amount of distraction caused by static distracted with that of moving distractors. The results are summarised in Figure 6.5.

With SVP, test subjects correctly identified 49.3% (SD = 24.09%, N = 48) of static targets in the scenario. With LSVP, test subjects correctly identified 49.7% (SD = 25.89%, N = 66) of static targets in the scenario. Statistical significance analysis using the t-test indicates that the hypothesis that LSVP is more successful than SMP for moving targets is not significant at a 5% confidence interval ($p=0.54>0.05$). **$H_{1.4}$ is rejected. SVP does not outperform LSVP for static targets.**

The conclusion from this result states that the moving distractions in our scenario did not distract the test subjects more with LSVP than when they were static in SVP. This can be related to the findings of Holcombe (2010) and Abrams (2003) – summarised in Chapter 4.4.1 – that conclude that unless the movement of an object is odd or unusual in some way it will not attract the subjects attention more than if it was static.

Figure 6.5: Target Identification Success Rate (in %) for each target. The green bar represents the percentage of correctly identified targets, whereas the red bar represents the percentage of targets that were not identified.

***H*$_{1.5}$:** Overall, LSVPT will obtain a higher target identification success rate than LSVP, using a high level of distractions.

In $H_{1.5}$ we expect that LSVPT will perform better than LSVP because the cameras are tilted in such a way as to reveal a larger area of the target. A profile view of the target will give better detail of their identity, subsequently they should be easier to identify.

With LSVP, test subjects correctly identified 55.3% (SD = 22.05%, N = 132) of all targets in the scenario. With LSVPT, test subjects correctly identified 61.3% (SD = 19.37%, N = 90) of all targets in the scenario. Statistical significance analysis using the t-test indicates that the hypothesis that LSVP is more successful than SMP for moving targets is significant at a 5% confidence interval ($p=0.026<0.05$). **$H_{1.5}$ is confirmed.**

Using 2 tilted cameras to provide an oblique view of the river banks (LSVPT) with the viewing mode of LSVP shows an improvement in target identification rate in comparison to simple LSVP.

This is caused by the test subjects obtaining more information regarding the targets they are observing. In LSVPT, it is easier to notice target clothing and characteristics.

Summary of subjective questionnaire results

The subjective questionnaire revealed some interesting facts about the opinions of the test subjects regarding the different presentation modes. Most interesting was the answer to the following questions:

1. How easy was it to perform the task?
2. How confident are you that you found all the targets?
3. Did you find the experiment enjoyable?

The results for questions 1 and 2 were combined to give a measure of the confidence with which the test subject performed the experiment Figure 6.6.

Whereas Figure 6.7 shows the average score given as the direct answer to question 3.

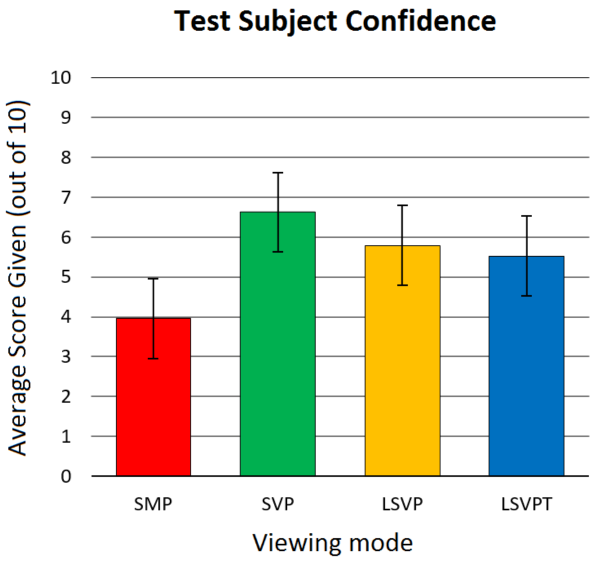

Figure 6.6: Evaluation of the test subject's confidence after performing the experiment. Scored from 0 to 10.

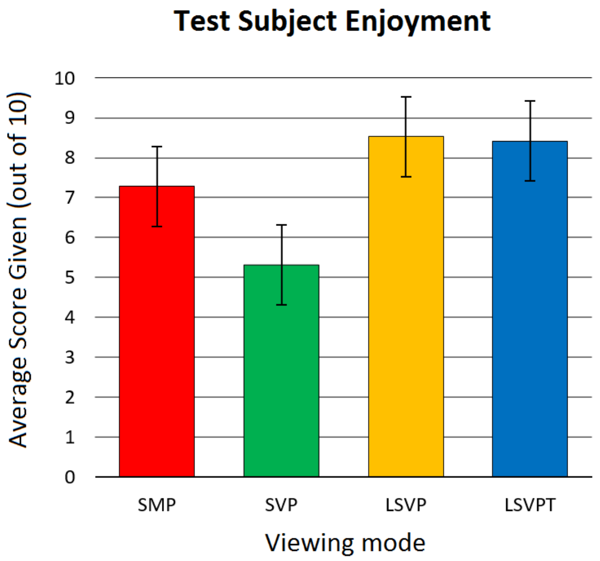

Figure 6.7: Evaluation of the test subject's enjoyment of the experiment. Scored from 0 to 10.

Evaluating test subjects' level of confidence. We find statistically significant result derived as before, showing the following:

1. SMP shows a lower confidence level than all the other mode. We suggest that this is caused by the lower number of identified targets as well as the absence of spatial context that was previously analysed.
2. LSVP and LSVPT show similar level of enjoyment that is insignificantly different. The two modes are very similar and therefore it is not surprising that they attained a similar score. Both are significantly higher than the score for SMP, the reason suggested being the presence of spatial context.

3. SVP shows a significantly higher confidence level relative to all the other methods. We suggest that this is caused by the provided spatial context, as well as the fact that the environment (distractions and feature) do not move in the still images provided by SVP, making it easier on the eyes for the observer.

Evaluating test subjects' level of confidence. We find statistically significant result derived as before, showing the following:

1. SVP shows a lower enjoyment level than all the other mode. We suggest that this is caused by the static frames which it provides. The absence of any movement makes the experiment dull for test subjects.
2. SMP shows a significantly higher enjoyment level in comparison to SVP. We suggest that this is caused by the presence of movement. It also shown a significantly lower level than observed with LSVP and LSVPT. We suggest this is caused by the lower confidence that was observed previously.
3. LSVP and LSVPT show similar level of enjoyment that is insignificantly different. The two modes are very similar and therefore it is not surprising that they attained a similar score. They obtain a significantly higher score than the other modes. The suggested reason is that movement is observed which entertains the subject, but at the same time the confidence level is kept at a decent level as previously observed.

6.3 Evaluation

Target Design

After analysis of the results and their conclusions, we tried to identify the causes of an intriguing anomaly that was observed. We expected to see trends in target identification such as, LSVP showing the best target identification rate for all moving targets. Contrary to our expectations, we observed that in Experiment 1, SVP performed significantly better than the other modes for moving targets 3 and 4, but worse for target 6. After careful evaluation we identify a flaw in target and viewing mode design that naturally puts SVP at an advantage for targets 3 and 4.

Target 3 is a child sitting in a boat that moves along the river, in the opposite direction to the camera. In SVP, the camera captures a snapshot on the terrain when the target is in the middle of the screen, resulting in 6 seconds (period between switches of frame) of static display of this target on the screen. In SMP, the camera and the target move in opposite directions, by calculating the speed of the camera and the boat, it is found that the target is only shown on the screen for 4 seconds. In LSVP, we are observing a frame (without the target) and when the target is about to appear on the screen from the top, the frame switches over to the next frame further up the river. The result being that now the target is located right in the middle of the screen and proceeds to move downwards with the boat's velocity, resulting in a display time of 3 seconds. Naturally, the more time the target spends on the screen, the more we increase the chance of detection by the test subject.

Similarly with Target 4, a child that runs from under a bridge and hides behind a bus stop. It takes the child 3 seconds to perform this movement during which he is visible, during the other 3 seconds he is either under the bridge or behind the bus stop where he cannot be seen. In SMP and SVP, we observe the full movement of the target resulting in 3 seconds during which it can be spotted by the test subject. In contract, during SVP, the snapshot is taken whilst the target is in plain sight, resulting in 6 seconds during which the target can be spotted. Additionally, even if target 6 is a moving target, it

This is a very clear explanation as to the causes of the results for targets 3 and 4. However, SVP was designed to take the snapshot at these times so that the targets would be seen. The probability that during SVP each one of targets 3 or 4 did not appear at all is ¼ and it so happened that in our scenario they both appeared, but this can put SVP at a disadvantage in another scenario. Ultimately, our scenario was not designed in the best way to draw conclusions, however it was realistic in the sense that in the real world, either of the 3 viewing modes can be at an advantage or a disadvantage depending on the behaviour of the target in this context. However, it must be noted that with SMP and LSVP there is no chance that a target is not shown at all on the screen for any amount of time.

Test subjects

As previously mentioned, it was not possible to control the age of test subjects efficiently given that we had limited time at the Festival to conduct our experiment. Younger and older individuals tended to score less than people in their prime age. Even though we tried to even out this number as much as possible, it might have affected the results.

Additionally, we found that not all test subjects read the instructions carefully, for this reason we included 2 example targets at the beginning of the scenario in order to make it clear what they were looking for. We attempted to design the example targets to be as generic as possible, but there exists a possibility that a test subject would take them too literally and only look for targets that look exactly the same. On the contrary, some test subjects might have taken then too vaguely making them less prepared for the experiment.

The conclusions of the subjective questionnaire show that test subjects found some viewing modes less enjoyable than other. This could have resulted in a loss of focus and motivation which would naturally lead to lower scores.

7 Experiment #2: Low Distraction

For Experiment 2 we use a modified scenario where the amount of distractors was reduced. Fewer cars, pedestrians and boats were featured. Additionally we increased the difficulty of the targets because they were now much easier to find given that there was less distraction.

The same Viewing Modes as in Experiment 1 were used:

1. Standard Moving Presentation
2. Serial Visual Presentation (SVP)
3. Live Serial Visual Presentation (LSVP)

7.1 Experimental Procedure

Experiment 2 was conducted in Bulgaria as we considered that there is good opportunity to find a decent number of test subjects which would be different from the ones in Experiment 1. Most subjects were friends, colleagues and acquaintances.

The same process as before is repeated. Test subjects are requested to take a seat on a desk with the apparatus set up. The display monitor used was slightly different but we maintained the aspect ratio. The same instructions are given, together with the examples at the beginning of the sequence. The results are recorded by the experiment supervisor.

A pilot study was conducted, similarly to Experiment 1. With the help of the first 10 test subjects we modified the difficulty of the targets so that the success rate would be closer to 50%, as before.

We then proceeded to collect a total of 48 surveys, 16 for each presentation mode.

7.2 Hypotheses and Results

***$H_{2.1}$*:** Overall, LSVP will obtain a higher target identification success rate than SMP and SVP, using a low level of distractions.

If $H_{2.1}$ is confirmed it will show that LSVP can perform better than SMP and SVP when the scenario features a low level of distractions.

Figure 7.1 shows the overall results for the 3 Viewing Modes in the scenario with low level of distractions. With SMP, test subjects correctly identified 51.3% (SD = 18.1%, N = 96) of all targets in the scenario. With SVP, subject correctly identified 67.5% (SD = 17.9%, N = 96) of all targets. With LSVP, subject correctly identified 68.1% (SD = 17.2%, N = 96) of all targets. Statistical significance analysis using the t-test indicates that the hypothesis that LSVP is superior to SMP is significant at a 5% confidence interval ($p=0.0002<0.05$). Statistical significance analysis using the t-test indicates that the hypothesis that LSVP is superior to SVP is not significant at a 5% confidence interval ($p=0.51>0.05$). **$H_{2.1}$ is rejected.**

LSVP cannot be identified as giving a significantly higher target identification success rate than SVP However, its result represents a significant improvement over SMP.

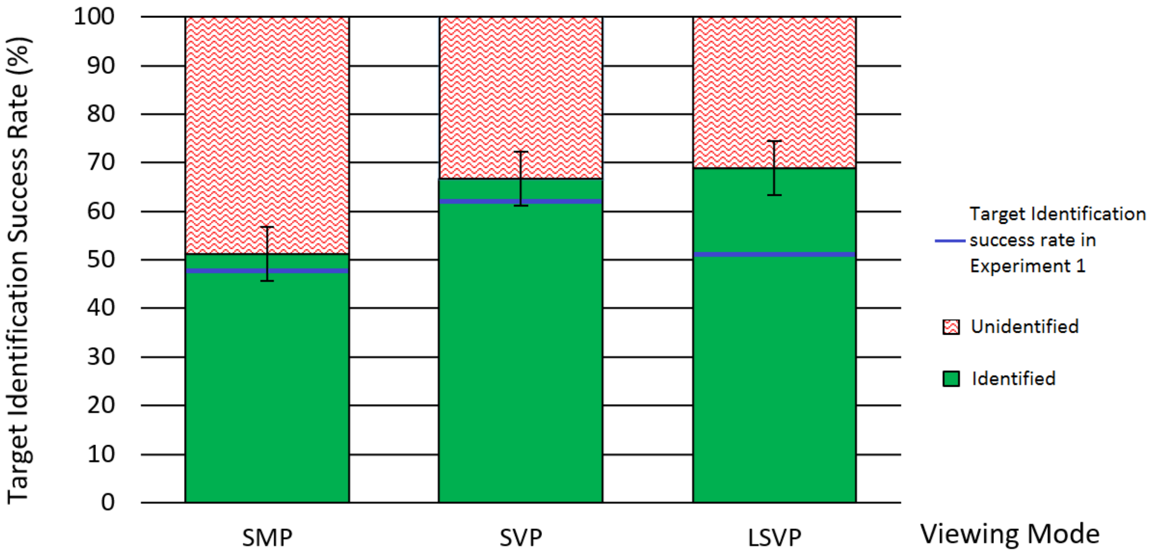

Figure 7.1: Summary of the overall target identification success rate (in %) for the 3 Viewing Modes tested in Experiment 2. The green bar represents the percentage of correctly identified targets, whereas the red bar represents the percentage of targets that were not identified. The black bar represents the standard error (at ±1.960).

***$H_{2.2}$*:** Overall, LSVP will show a better improvement in target identification rate in comparison to SMP and SVP between the scenario with a high level of distractors (Experiment 1) and the scenario with low level of distractors (Experiment 2).

If LSVP does not come out straight away as the better Viewing Mode to use in our scenario with a low level of distractors ($H_{2.1}$ rejected), we can further investigate whether a significantly larger improvement is observed for LSVP comparing to the other modes between the 2 experiments. A comparative improvement will reveal potential for LSVP to perform even better in a scenario with even less distractors such as Wilderness SAR.

In order to identify whether there is potential for LSVP to outperform the other Viewing Modes given the right circumstances, we must observe the improvement that is shown by LSVP in the second experiment. We compare the improvement shown by SMP and SVP to the improvement shown by LSVP between the two experiments (Figure 4.2). In order to calculate the t-score, we must first combine the standard deviations of the 2 data sets (from experiment 1 and experiment 2) that were combined to obtain the improvement rate (Ballew, 2005). With SMP, the overall percentage change between experiment 1 and 2 is 6.97% (SD = 4.91%, N = 180). With SVP, the overall percentage change is 6.5% (SD = 4.56%, N = 192). With LSVP, the overall percentage change is +26.4% (SD = 4.75%, N = 228). Statistical significance analysis using the t-test indicates that the hypothesis that LSVP improves more than SMP is significant at a 5% confidence interval ($p=0.0001<0.05$).Statistical significance analysis using the t-test indicates that the hypothesis that LSVP improves more than SVP is significant at a 5% confidence interval ($p=0.0001<0.05$). **$H_{2.2}$ is confirmed.**

This result indicates that there is a significant difference between the improvement of LSVP, and those of SMP and SVP. LSVP therefore benefits from an advantage when less distractions are present.

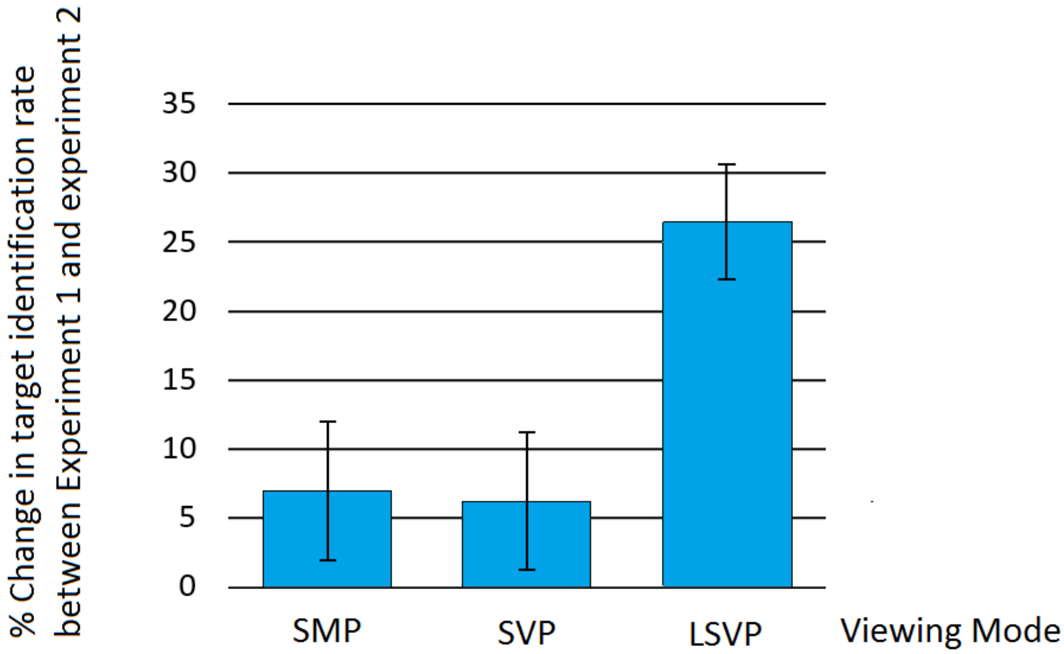

Figure 7.2: % Change in target identification rate between Experiment 1 (high distraction) and Experiment 2 (low distraction. The black bar represents the standard error (at ±1.960).

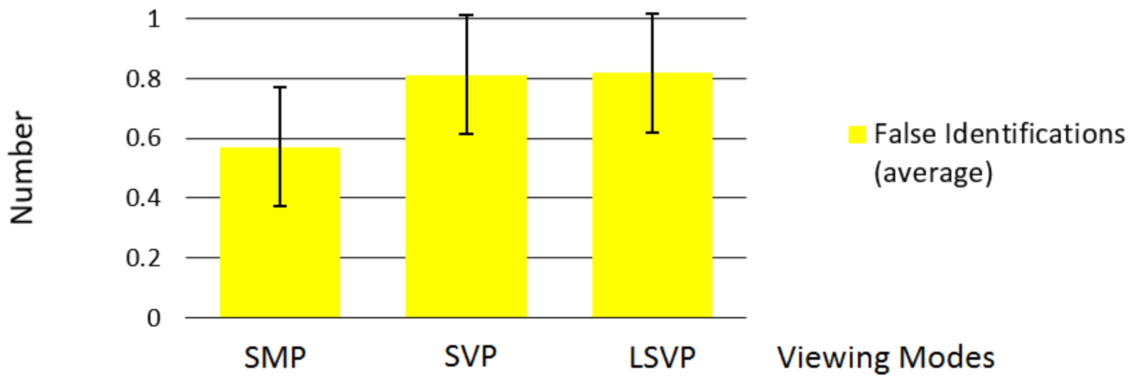

Figure 7.3: Average number of false identifications (False Positives) for the 3 viewing modes. The black bar represents the standard error (at ±1.960).

***$H_{2.3}$*:** LSVP will perform better than SVP and SMP for targets categorised as "Moving", using a low level of distractions.

If this hypothesis is confirmed, we can additionally state the LSVP is superior to SVP and SMP when looking for a target that is expected to be performing a movement that will help identify it, such as waving, or running.

In Figure 7.4 we observe the target identification success rate for moving targets (bottom). Statistical significance exists to conclude that LSVP performs better than MSP and SVP for moving targets 4 and 6. However it performs worse than SVP for target 3. As previously explained in Chapter 6.5 – Experiment 1 Evaluation, there is a flaw in the design of Target #3, we therefore exclude this target for the analysis of this hypothesis. With SMP, test subjects correctly identified 54% (SD = 17.1%, N = 48) of moving targets in the scenario. With SVP, test subjects correctly identified 58.7% (SD = 16.92%, N = 48) of moving targets in the scenario. With LSVP, test subjects correctly identified 87.3% (SD = 17.92%, N = 48) of moving targets in the scenario. Statistical significance analysis using the t-test indicates that the hypothesis that LSVP is more successful than SMP for moving targets is significant at a 5% confidence interval ($p=0.0006<0.05$). Statistical significance analysis using the t-test indicates that the hypothesis that LSVP is more successful than SVP for moving targets is significant at a 5% confidence interval ($p=0.0005<0.05$). **$H_{2.3}$ is confirmed.**

Leading us to a firm conclusion that when there are less distractors in the scenario, LSVP outperforms the other modes for moving targets. This confirms our theory from Chapter 4.4.1 that the more distractions there are in a scenario, the lower the chance for the test subject to identify moving targets performing an odd or unusual movement.

Figure 7.4: Target Identification Success Rate (in %) for each target. The green bar represents the percentage of correctly identified targets, whereas the red bar represents the percentage of targets that were not identified.

7.3 Evaluation

After analysing the results from Experiment 2, we found that the age range of test subjects was much narrower than it was for Experiment 1. The reason being that friends, colleagues and acquaintances of the experiment supervisor and recruiter of test subjects were appointed as test subjects. As previously observed, people in their prime (17-47) performed best. Given that 87.5% of test subjects in Experiment 2 were in that age range as opposed to 71% in Experiment 1, it can be argued that this factor caused results to be higher in Experiment 2.

Additionally, a large proportion of test subjects were IT specialists and Engineers who are typically accustomed to focusing on a computer display for large portions of the day and are more likely to be able to focus better and navigate the screen with their eyes more efficiently.

8 Overall Evaluation

After previously evaluating the two experiments separately, this section will relate the project requirements and analyse our accomplishment unrelated to the specific experiments themselves.

LSVP and its practical implications

As was previously discussed in Chapter 4.4, the implementation of LSVP requires two UAVs with a specifically programmed flying pattern in order to provide the video to the observer. This implies the need of additional funds by SAR and therefore might not make it the most efficient solution to investigate in our research. In retrospect, it might have been a good idea to also test LSVPR – LSVP with only one UAV where the camera rotates – with the chance that it would achieve a similar result but be more cost efficient.

Testing Platform and Scenario Realism

The choice of game engine development software was shown to be very successful and provided us with a lot of flexibility and efficiency for designing the scenarios. However, after analysing the completed scenarios, we find that there is more work to be done in terms of environment realism. We identify the key points that should be improved as the following:

- Pedestrians were adventure character models because they were the easiest to animate, these are not directly relevant to the style of the city used for the scenario
- Pedestrians should employ different walking animations as well as more sophisticated walking patterns (not just in a straight line)
- More sophisticated assets should be used in terms of detail in order to provide a clear video to the target spotter
- The building and the style of the city, in general, should be redesigned to reflect a more contemporary style where the test subjects would feel “at home”
- The river resembled a channel rather than a real river as it was straight and looked artificial

Targets

During the statistical analysis faze, we found that if we included more targets to our scenarios it would increase the credibility of the experiment. Additionally, reflection on the user experiment has shown that test subjects believed that there were a larger number of targets and therefore needed encouragement to continue to the end of the scenario. Considering that including more targets in simulation would have been a quick and efficient process, we regret to not have done it.

Eye – Tracking

During the experiment we observed that with SMP, the most commonly used technique by test subjects was to scan the top of the screen left and right. In contract, with SVP and LSVP they started from a top corner of the screen, going down, across horizontally, and back up. This also contradicts the findings of Mardell (2013). The suggested cause being Urban versus Wilderness SAR. If we had implemented eye – tracking and recorded this occurrence better, we would be able to draw important conclusions regarding the best visual techniques to be used in SAR depending on the scene.

Touch – Screen

We did not have enough time on our hands to implement an automated method to track target identification. Most importantly recording the exact distractions that were identified as targets but were not. This would allow us to analyse which features were most commonly mistaken for targets and provide this information to improve training for SAR spotters.

9 Overall Conclusions

The project involved many decisions and design choices as it progressed through the stages. The first most important choice was to use the game development software to create the simulation for SAR scenarios. This decision was very successful and we benefitted from many advantages in the long run. Most notably it was efficient for making changes and modifications resulting in a very beneficial user experiment where we had the opportunity to test a wide variety of modifications with a large sample of test subjects.

Our participation at the Imperial College Festival was of great benefit to the project, apart from the social outcomes, we obtained a decent set of data for Experiment 1 and drew important conclusions regarding Visual Search. Thanks to the success at the Festival, we were able to design a second experiment where we had the chance to investigate the level of distractions in the scenario.

Three parts can be identified as the most engaging, clever and challenging of the project. Firstly, implementing the simulator was the most engaging, this task took the longest time but brought a lot of learning outcomes with it. I gained a lot of experience in working with game development software as well as programming in C#, the most widely used language in modern simulator and game development. The natural evolution of this task went into the design of the user experiments, this was the cleverest part of the project. Identifying the best presentation modes to verify, target design and implementation of the environment to suit the user experiment were the key points for this step. Conducting the experiments and analysing the data was the most challenging part, it required analytic thinking and decision making in order to decide in which way to progress our project for Experiment 2. Most difficulties were overcome with hard work and a lot of support from my supervisors.

In general, the experiments and the analysis of their results were aimed at comparing the 4 video presentations modes – SMP, SVP, LSVP and LSVPT. The aim was to come to a conclusion as to which one will facilitate most the finding targets given 2 Urban SAR scenarios – one with a high level of distraction, and the other with a low level of distractions.

The findings show that in the presence of many distractions in the scenario, SVP facilitates the finding of targets because it provides spatial context to the video frame observed. In this scenario, the newly invented viewing mode LSVP did not perform so well, the reasons being that the distractions took away the attention of the test subject who was not able to spot the movement of the targets. In brief, the presence of many distractions cancelled out the benefits of LSVP. For this reason, a second experiment was conducted where we decreased the level of distractions. Experiment 2 was not definitive on concluding LSVP as a superior mode in the scenario. However, LSVP showed a significantly better target identification success rate for moving targets. Additionally, its performance showed the highest improvement between the scenario with a high level of distractions (Experiment 1), and the one with a low level (Experiment 2).

These conclusions clearly show that the advantage for LSVP lies within scenarios with low level of distractions where targets are expected to be moving. Whereas SVP performs better in scenarios with many distractions (moving and static), where the movement in the target might not be the main characteristic to look for.

Additionally we find that providing an oblique view through LSVPT obtains better target identification rates due to the spotter being revealed more information about the targets characteristics such as clothing and height.

10 Further Work

Combination of SVP and LSVP

As observed during the two experiments, both SVP and LSVP have advantages over SMP and offer an increased chance for target identification. As observed by Dr Witkowski, there is potential for an additional viewing mode that combines SVP and LSVP and only requires 1 UAV in comparison to LSVP where 2 are required, however a higher UAV flight speed is required.

The principle is best explained by an example:

1. Set T = 6 seconds to be the time that each frame of the terrain is displayed
2. Set the distance between 2 frames to be 20m
3. At t = 0, the UAV is situated and filming over Frame 1 for 3 seconds, performing LSVP
4. At t = 3, the UAV captures a still image that is projected as SVP, meanwhile, during the next available 3 seconds (T-t) it flies over to Frame 2, 20m away, with an average speed of 6.67 m/s
5. At t = 6, the UAV is over Frame 2 and ready to performs step 3 and 4 for again

Because this method will benefit from the advantages of both SVP and LSVP, it is suggested that it has the potential to be used for high and low distractions scenarios with minimum losses.

Visual Search and Attention – Eye tracking and controllers – Spotter training

As previously explained in Chapter 8 – Evaluation, it was suggested that implementing eye – tracking when performing the experiment would have provided us with valuable information regarding the search techniques employed by the test subjects. Subsequently, suggestions on the most successful techniques could be made.

In addition to eye tracking, controllers for the camera in the simulation can be provided to test subjects. This will allow the investigation of "Where the spotter tends to look for the targets?" or "Where is the attention of the spotter naturally attracted to?"

In general, even though spotter training has previously been investigated in the field, it might be worthwhile to investigate further using the modern methods and simulation that was created.

Wilderness and Levels of Distractions

As a conclusion from our experiments, we found that the success of different viewing modes (especially LSVP) greatly depends on the level of distractions in the scenario. It will therefore be intriguing to investigate how LSVP, and other viewing modes will perform in a Wilderness SAR scenario.

Live Visual Serial Presentation Tilted (LSVPT)

Following the result from Experiment 1, and the significant improvement that was observed with LSVPT, we consider it worthwhile to further investigate the modification of the viewpoints from the UAV camera and conduct user experiments to conclude how they perform given the different scenarios. In this context, UAV height, number of cameras, camera angles should be further researched.

Virtual Reality

Virtual Reality (VR), or in other words, a 3D headset attached in front of the eyes, has been gaining popularity during the recent year. An investigation into whether VR can help in target identification by submerging the test subject into the scenario can lead to interesting discovery using innovative technology.

Scenario Simulation

As outlines in Chapter 9, the existing simulation can be reused and reworked to represent a more realistic scenario.

11 Unity 3D Project User Guide

Together with the project we offer all the developed software in Unity 3D together with all the purchased assets and the scenes that were used for the experiment.

The Unity 3D project comes as compressed folder provided on a CD to our supervisors and the department as well as on Google Drive under the following link:

https://drive.google.com/open?id=0BxlitMXs6UNQRi1BSE1PdTR5d1E

The work was completed on Unity v5.2.2f1 (64-bit). All one needs to do is open the project using Unity. The software is very flexible, even if you use a newer version to open it, it should be able to adapt the project and allow you to make full use of it.

Within the compressed folder we provide a ReadMe text file with the account in the Unity 3D Asset store which includes all the assets that were purchased. The rest of the assets (the free ones) will be found within the Assets folder in the project.

12 Bibliography

Abrams, R. A. (2003). Motion onset captures attention. *Psychol Sci.*, (5):427-32.

Ballew, P. (2005). *Combining Two Distributions*. Retrieved from pballew: http://pballew.net/combdis.htm

Beena Khurana, E. K. (1987). Shared attentional control of smooth eye movement and perception. *Vision Research 27(9)*, 1603–1618.

Burr, D. C. (1982). Contrast Sensitivity at HighVelocities. *Vision Research 22*, 84-479.

Carrasco, M. D. (1995). The Eccentricity Effect: Target Eccentricity Affects Performance on Conjunction Searches. *Perception & Psychophysics 57 (8)*, 61-1241.

Chan, A. H. (1996). Foveal Acuity, Peripheral Acuity and Search Performance. *A Review. International Journal of Industrial Ergonomics 18 (2-3)*, 9-113.

Croft, J. L. (2007). Gaze Behavior of Spotters During an Air-to-Ground Search. *Human Factors: The Journal of the Human Factors and Ergonomics Society 49(4)*, 8-671.

Goodrich, M. A. (2008). Supporting Wilderness Search and Rescue using a Camera-Equipped Mini UAV. *Journal of Field Robotics*, 89-109.

Holcombe, C. J. (2010). Unexpected changes in direction of motion attract attention. *Attention, Perception, & Psychophysics 72 (8)*, 2087-2095.

Humphreys, G. W. (1993). SEarch via Recursive Rejection (SERR): A Connectionist Model of Visual Search. *Cognitive Psychology, no. 25*, 43–110.

Kowler, E. (2011). Eye Movements: the Past 25 Years. *Vision Research 51 (13)*, 83-1457.

Mardell, J. P. (2013). Assisting Search and Rescue through Visual Attention. PhD Thesis, Department of Electrical and Electronic Engineering, Imperial College London.

Mardell, R. S. (2012). A Comparison of Image Inspection Modes for a Visual Search and Rescue Task. *Behaviour & Information Technology 33(9)*.

Nakayama, K. a. (1986). Serial and Parallel Processing of Visual Feature Conjunctions. *Nature 320 (6059)*, 5-264.

RAF, M. A. (2011). Working with Search and Rescue Helicopters. *Royal Air Force, SAR FORCE: VERSION 1.0*. London, United Kingdom: Royal Air Force Search and Rescue.

Rudol, P., & Doherty, P. (2008). *Human Body Detection and Geolocalization for UAV Search and Rescue Missions Using Color and Thermal Imagery*. Linköping: Department of Computer and Information Science, Linköping University.

Stager, P. a. (1975). *Eye-Movements and Related Performance in SAR Visual.* Defence and Civil Institute of Environmental Medicine.

Tatler, B. W. (2003). On Nystagmus, Saccades, and Fixations. *Perception 32 (2)*, 84-167.

Tinman, R. (Director). (2014). *RC Search and Rescue* [Motion Picture].

Treisman, A. M. (1980). A Feature-Integration Theory of Attention. *Cognitive Psychology 136*, 97–136.

Treisman, A. M. (1980). A Feature-Integration Theory of Attention. *Cognitive Psychology 136*, 97–136.

UK Emergency. (2016, 06 05). *UK Emergency Vehicles*. Retrieved from UK Emergency: http://www.ukemergency.co.uk/category/police/police-air-support/

Verghese, P. (2001). Visual Search and Attention: a Signal Detection Theory Approach. *Neuron 31*.

Watson, H. K. (2013). *Seventh Framework Programme – Security Research; Deliverable 1.2: Report on search and rescue actions.* London: Cosmic.

Wolfe, J. M. (2004). What Attributes Guide the Deployment of Visual Attention and How Do They Do It? *Nature Reviews Neuroscience 5 (6):*, 495–501.

Wolfe, J. M. (2007). Low Target Prevalence is a Stubborn Source of Errors in Visual Search Tasks. *Journal of Experimental Psychology 136 (4)*, 38-623.

Woodman, G. F. (2003). Serial Deployment of Attention During Visual Search. *Journal of Experimental Psychology: Human Perception and Performance 29 (1)*, 38-121.

13 Appendices

Appendix A – Test Subject Instructions

- A drone with a video camera flies over a river. You will be observing the video from its camera.

- Your task is to find 6 lost school children along the river. All the children are wearing a red top and blue trousers, see the example.

- Whenever you think you see one of the children, point to that child on the screen. Don’t forget to look along the river, as well as the sidewalks.

- The sequence will last about 3 minutes, please watch it to the end.
- Say START when you are ready.

Appendix B – Participant Questionnaire

Age group:

☐ 3-17 ☐ 18-27 ☐ 28-37 ☐ 38-47 ☐ 48-57 ☐ 58+

Gender:

☐ Male ☐ Female

Eyesight:

☐ Good or Corrected (with glasses) ☐ Impaired or not wearing prescribed glasses ☐ Colour-blind

Do you study, or have you studied at the college (including research)?

☐ Yes ☐ No

You have watched the sequence. Please give your opinions of it, by circling a number for each question.

How easy was it to perform the task?

not at all 0 1 2 3 4 5 6 7 8 9 10 *extremely*

How confident are you that you found all the targets?

not at all 0 1 2 3 4 5 6 7 8 9 10 *extremely*

How focused were you on the task?

not at all 0 1 2 3 4 5 6 7 8 9 10 *extremely*

How realistic did you find the sequence to be, in terms of aesthetics and behaviours?

not at all 0 1 2 3 4 5 6 7 8 9 10 *extremely*

What do you think of the speed with which the camera moved along the river?

very slow 0 1 2 3 4 5 6 7 8 9 10 *very fast*

Did you find the experiment enjoyable?

not at all	0	1	2	3	4	5	6	7	8	9	10	*extremely*

Additional Comments:

....

....

....

....

....

Appendix C – Recording the Results