Emergency Urban Search

Nikola Dourtchev, Mark Witkowski, James Mardell and Robert Spence
Department of Electrical and Electronic Engineering
Imperial College London
United Kingdom
Email: ndourtchev@gmail.com

Abstract

The use of a video camera to support search immediately raises the question "for target location, what is the most effective way of presenting video camera output to a human observer ('spotter')". We examine three presentation modes: (a) unprocessed video output, (b) 'static visual presentation' (SVP) in which a series of static views of the search area can be examined in turn while keeping up with drone movement, and (c), a novel mode called 'Live SVP' in which the locations sequentially captured by a camera are presented discretely in real time, thereby preserving any movement such as a person waving to attract attention.

The dynamics of aerial video were modelled using game development software. The resulting videos were used to support realistic search exercises using human participants. The task attempted by each participant was the identification of lost school children in the simulated environment, using one of the video presentations described above.

It was found that the new LSVP viewing mode is superior in those tested for moving targets in a low distraction environment. Another principal finding was that the density of distractors (i.e., non-target objects) had a significant influence on the success of target identification.

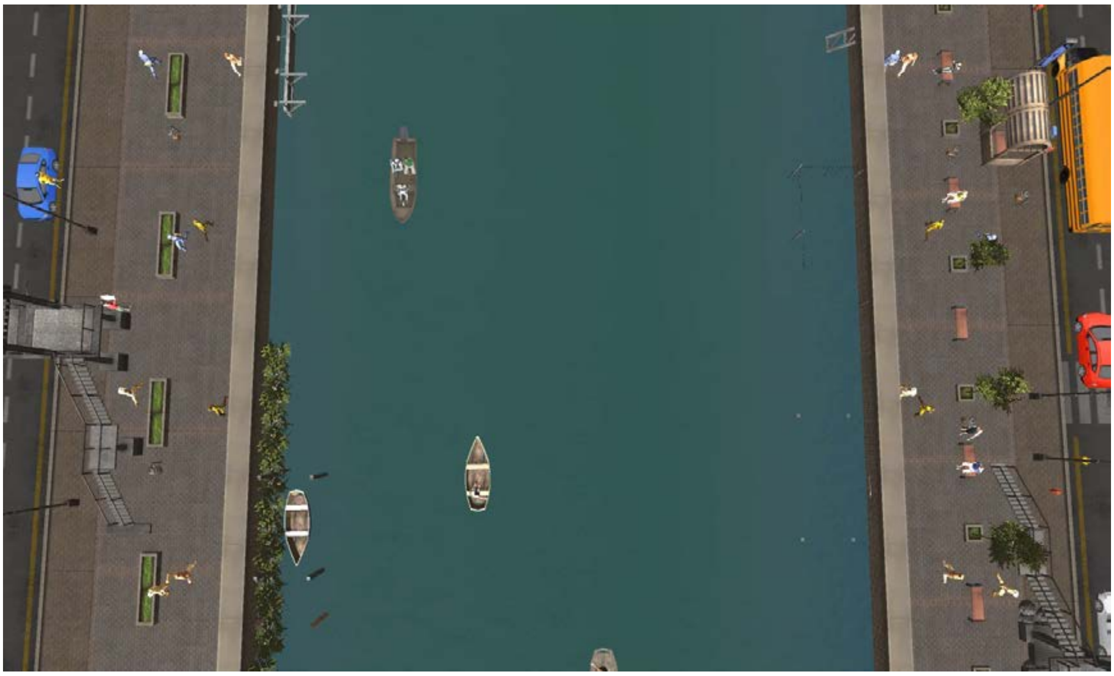
KEY WORDS: Search, Rescue, Drone, Video, Target, Presentation

Introduction

There are many scenarios within an urban environment where some type of 'target' must be located as quickly as possible. The target might be a dementia patient who has wandered unobserved, the victim of a boating accident, an escaping criminal, a person apparently intent on suicide or some artefacts associated with the person being sought. Typically a search vehicle (e.g., boat or automobile) and driver are called out, together with one or more human spotters who, with or without optical aids, conduct a visual search along a route determined by received intelligence and accumulated experience.

The likelihood of a successful search can potentially be enhanced in at least two ways. An obvious one is to employ drones carrying downward facing video cameras, since the flexibility of their positioning can provide a wide variety of views not always available to a conventional human spotter. A second way is to process and present the raw camera output in such a way as to enhance the likelihood that a spotter will identify a target. The latter approach is investigated in the study reported here.

Figure 1: A typical simulated output of a drone-borne video camera



Simulation

To perform the intended formal investigation it is necessary to control the experimental parameters. For financial and logistical reasons that is infeasible within a real-life search and rescue scenario, so simulation was employed, using game development software to model the overflown terrain as well as the dynamics of drone movement and the generation of video camera output. The resulting views of terrain provided by processed video camera output were of a quality more than sufficient to carry out realistic tests using human participants as spotters. An example of simulated processed video camera output is shown in Figure 1, which depicts a canal and part of its shoreline. Both the canal and its shoreline are populated with commonplace objects, some of which are the targets being sought.

Typical objects, both human and non-human, are illustrated in Figure 2.

Figure 2: Some objects populating the canal and its shoreline

Related work

The enhancement of urban search and rescue calls for attention to be paid to a very wide range of issues spanning many disciplines. Of topical interest is the use of a drone, the output of whose downward facing camera is transmitted to a base station where trained human spotters attempt to identify a missing person or related cues as to their location (Simerson 2004, Goodrich et al. 2008; Doherty & Rudol, 2007). The resulting flexibility of search, as well as the issues demanding investigation, increases when multiple drones (swarms) are deployed: constraints are imposed by the environment (e.g., obstacles, weather conditions) (Wahate & Trigoni, 2010) and by the topography, the latter also impacting upon communication between swarms and ground control (Wahate et al, 2009).

There are many essentially human issues to be addressed. One is that of training: the typical success rate of untrained personnel can be as low as 30% (Croft, 2007), and research has shown that a trained "spotter" can only provide an improvement of about 10% (Stager, 1975). Nevertheless, even expert spotters can suffer the detrimental effects of inattentive blindness (Drew et al, 2013). Issues related to human cognition and visual perception can draw upon decades of research: see, for example Wolfe & Pashler (1998), Wolfe (2004, 2007) and Ware (2012). In these and other reports, the significance of preattentive processing (Healey & Enns, 2012) and distractors is apparent (Wolfe et al, 2002; Wolfe & Gray, 2007). Recently, investigations of visual search have been set within the specific context of search and rescue (Adams et al, 2009; Mardell et al, 2013) and have focussed upon the influence of human cognition and visual perception. For wilderness search and rescue Mardell et al (2013) investigated how the perceptual ability of a human spotter can be supported by the appropriate processing and presentation of the images obtained by a drone's camera: they discovered that a sequence of static views led to target identification performance significantly better than unprocessed video output. The general focus of the present paper is the same, but specifically directed to urban search and rescue and the proposal and evaluation of a new method of image processing of potential value when targets are moving.

Processed video output

In the context of wilderness search and rescue, Mardell et al (2011, 2013) have evaluated a number of ways in which the output of a downward pointing video camera might usefully be processed before presentation to a spotter. Six possibilities were evaluated of which two are directly relevant to the present study:

> Standard Moving Presentation (SMP): No processing takes place and camera output is displayed directly to the spotter.
>
> Serial Visual Presentation (SVP): here, the spotter is presented with a sequence of static images of the terrain captured by the camera as the drone moves along a path. Each image is presented to the spotter for a specific amount of time until it is replaced with the next one, though at a rate that 'keeps up' with the drone. In the context of wilderness search and rescue it was found (Mardell et al, 2013) that, statistically, SVP leads to significantly better target identification than does SMP.

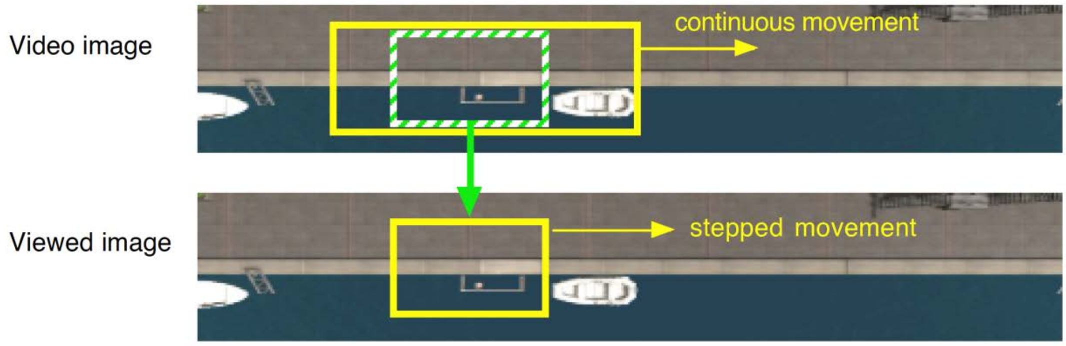
Despite the identified advantage of SVP over SMP, a remaining problem associated with the SVP viewing mode (but not with the SMP mode) is its inherently static nature. For example, if a missing person were to be waving in order to attract attention, this basically dynamic aspect of the captured image would be hidden from view in the SVP mode.

The need for a video processing mode, which retains the established advantage of SVP but allows a spotter to see temporal changes within a fixed area of terrain, prompted the investigation reported here. At the same time, it was decided to focus on urban search and rescue in view of its frequent daily occurrence and the typically higher density of distractors (i.e., non-targets).

A new viewing mode

Partly suggested by the work of Holcombe (2010) and Abrams (2003), a new video processing technique (Live Serial Visual Presentation - LSVP) was proposed that would support a viewing mode in which temporal changes are preserved. The principle is illustrated diagrammatically in Figure 3. The upper part shows (yellow box) the continuous moving image captured by the on-board video camera. Part (hashed) of that image corresponding to a fixed area of terrain is then presented (yellow box in the lower part of the figure) for viewing by a spotter, inherently retaining any temporal changes occurring in the original video image.

Figure 3: Schematic illustration of the basis of the tested Live Serial Visual Presentation ***(LSVP).***



Hypotheses

The principal aim of this experiment was to test three hypotheses. The first goal was to see if the results obtained by Mardell et al (2013) for wilderness search and rescue are equally relevant for search in the very different urban environment and with a high density of distractors:

H1: SVP will lead to a higher target identification success rate than SMP.

The second hypothesis concerns the effectiveness of the newly introduced Live SVP viewing mode. It was expected that LSVP would combine the advantage of a static view with the live movement associated with SMP:

H2: Overall, LSVP will lead to a higher target identification success rate than SMP and SVP.

A third hypothesis refers to the anticipated unique advantage of LSVP over SVP and SMP for targets that are moving:

H3: LSVP will perform better than SVP and SMP for moving targets

Experimental design

The scenario adopted for the experimental investigation comprised a canal and its immediate surrounding area. It was populated with people and artefacts (e.g., cars, boats and lighting fixtures) to constitute a representative environment in which an urban search and rescue might take place. Figure 1 illustrates an overall view of the search area, and Figure 2 illustrates various objects that populated it. In the reported experiments, six of those objects (all people) were designed to be targets, the remaining objects (including people) being considered to be distractors.

The task that was presented to participants taking part in the experiment was that of identifying six children who had been lost during a visit to the canal area. The lost students were characterized by their apparel, comprising a red top and blue trousers. Human distractions were introduced into the environment, wearing a red top with trousers that were not blue, and vice versa.

To allow study of the influence of different viewing modes, each of the six missing children constituted one of the following types of target:

Static target:

Sitting in the sidewalk area; standing in the sidewalk area; and within the river area on a static floating object

Moving target:

Running and hiding on the sidewalk (behind a bus stop); sitting in a moving object (e.g., boat); and standing and waving on the sidewalk.

Unity 3D" game engine software (http://www.unity3d.com/) was employed to create a model of the simulated environment (e.g., canal, sidewalk and populated objects) using items provided in the "Town Constructor Pack".

Pedestrians were modelled using the "Adventure Character Set" from the Unity 3D Asset Shop and their clothes created to show a range of colours. Their movement paths were defined along four virtual lanes, two forwards and two backwards. Figure 4 shows the six targets used in the experiment: they fall within the six categories defined above. Boats and cars were modelled and introduced into the scenario in a variety of ways. Twenty variations of boat type and number of passengers were introduced, and seven different models of car moved along the streets.

Figure 4: the six types of target: Top left (sitting); Top middle (standing); Top right (sitting in moving boat); Bottom left (running); Bottom middle (sitting); Bottom right (waving)

The simulated video output was chosen to be compatible with a drone flying at an altitude of 40 meters and moving at 6 meters per second.

Experimental procedure

A total of 56 participants took part in the first experiment. All were visiting a two-day university festival and volunteered their participation upon noticing an exhibit related to human vision. Their ages ranged from under ten to over seventy. In view of this age distribution the four viewing modes were, as far as possible, spread roughly equally over that age range.

Each participant was first introduced to the task by the experimenter reading from a script. They were then provided with a set of on-screen instructions with illustrations, stating:

"A drone with a video camera flies over a canal.
You will be observing the output from the drone's camera.
Your task is to find six lost school children along the canal.
All the children are wearing a red top and blue trousers.
Whenever you think you see one of the children point to that child on the screen.
The sequence will last about 3 minutes – please watch it to the end.
Say START when you are ready".

To familiarize each participant with the task and the environment, two example targets were introduced in the first part of the scene and pointed out by the experimenter: thereafter, after the drone had passed over a bridge that blocked the view, the participant was given no further assistance with target location.

Since the scenario was the same for each viewing mode, each participant was tested with only one viewing mode in order to avoid any learning effects. Each test lasted 3 minutes.

Every time the participant pointed to the screen the experimenter recorded the location and noted whether a target had been correctly or incorrectly located (i.e, whether the identification was a True Positive or a False Positive). At the conclusion of the task the participant was asked to complete a questionnaire concerning age range, gender and vision as well as subjective views concerning their experience. They were thanked for their participation.

Experience with the first experiment (see Experiment 1 below) and the pilots preceding it suggested that the relative advantages of the three viewing modes might be sensitive to distractor density, so a second experiment (see EXPERIMENT 2 below) was carried out to investigate this issue.

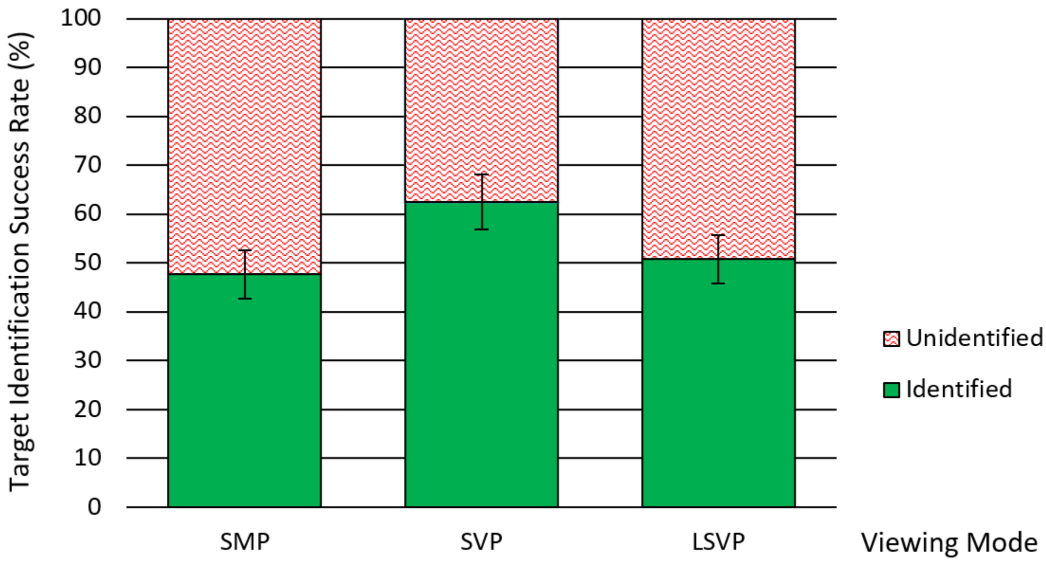
Experiment 1 (high distractor density)

Results

Target identification

The result of the first experimental investigation is shown in Figure 5. With SMP, participants achieved an average identification success of 47.6 % (SD = 18.3%, N=84), compared with 62.5% (SD = 22.4%, N=96) for the SVP mode. Statistical analysis using the t-test indicates that hypothesis H1 is confirmed at a 5% confidence level ($p=0.003 < 0.05$) in agreement with the findings of Mardell et al (2013) for wilderness search and rescue.

Figure 5: Result of Experiment 1 showing the target identification success rate (green) for each of the three viewing modes.



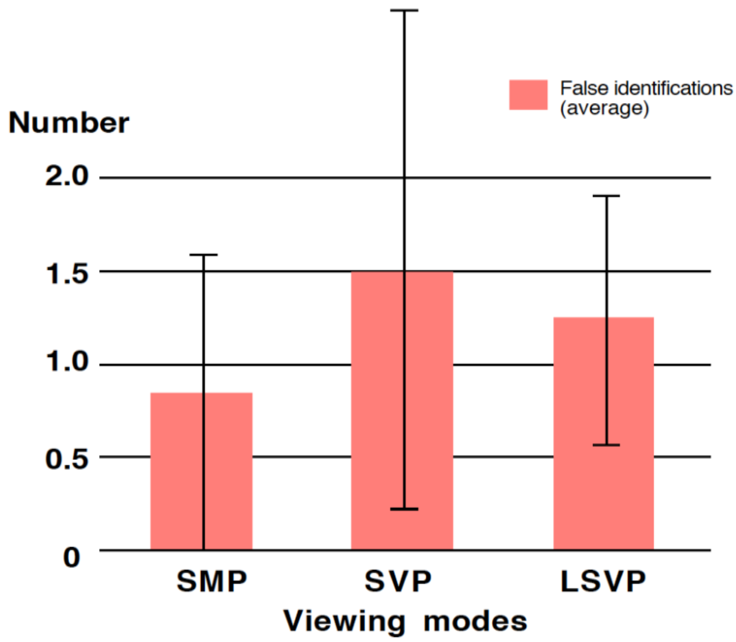
With regard to the LSVP viewing mode, the result shown in Figure 5 shows that participants correctly identified 50.7% (SD =26.2%, N=144) of targets. Statistical analysis using the T-test indicates that the first part of hypothesis H2 – that LSVP is more successful than SMP – is not significant at the 5% confidence level. Additionally, the analysis established that the second part – that LSVP is more successful than SVP – is also invalid at the 5% confidence level.

To summarise, H1 (that SVP will lead to a higher target identification success rate than SMP) is confirmed and H2 (that, overall, LSVP will lead to a higher target identification success rate than SMP and SVP) is rejected.

False positives

Also of interest is the occurrence of False Positives, where participants identified a target where one did not exist, potentially triggering wasted search effort. Figure 6 shows the false positive averages for the three modes investigated. Note that the SVP mode, which leads to more effective identification than SMP, is nevertheless associated with a significantly higher False Positive performance (significant at the 5% confidence interval ($p=0.006<0.05$).

Figure 6: The average number of false target identifications for the three viewing modes.



Moving targets

The new LSVP mode needs further consideration, especially in view of its expected ability to preserve movement when compared with SVP. We recall that, of the six targets, three were moving. Thus, with the LSVP viewing mode those targets are still moving, whereas in SVP mode none are moving. So it is appropriate to separate the results for the three moving targets from those for the static targets when viewed in SMP, SVP and LSVP modes, as presented in Figure 7.

Over all moving targets (targets 3, 4 and 6), participants correctly identified 48.8% (SD=22.91%, N=42)) with the SMP mode, 53.3% (SD=21.32% N=48) with SVP and 51.1% (SD=22.32%, N=66) with LSVP. Statistical analysis using the T-test indicates that 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), and that the hypothesis that LSVP is more successful than SVP is, again, not significant at the 5% confidence level ((p=0.99>0.05). Thus, there is no evidence that LSVP outperforms SMP or SVP for moving targets and hypothesis H3 (that LSVP will perform better than SVP and SMP for moving targets) is rejected for the conditions tested.

Figure 7: Target identification success rate (green) for each of the six targets, three static and three moving.

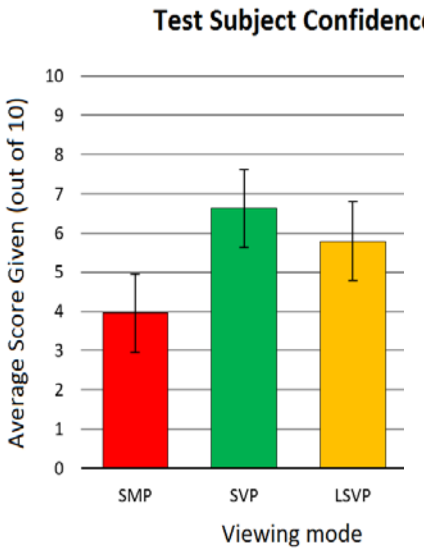
Static targets

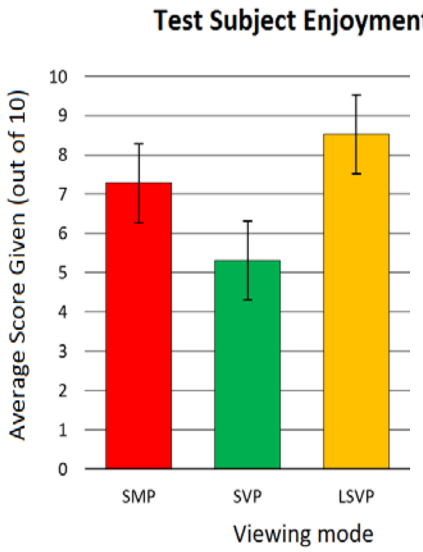
A related suggestion is that SVP will outperform LSVP for static targets. It can be postulated that moving distractions distract more than static distractions. Thus, the only change between the two tested cases is that in SVP the distractors will be static, whereas in LSVP they will be moving. All six targets were still presented, but identification success only recorded for the static targets. With SVP, participants correctly identified 49.3% (SD=24.09%, N=48) of static targets, compared with LSVP where correct identification occurred for 49.7% of static targets (SD=25.89%, N=66). Statistical analysis showed that the hypothesis that LSVP performs better than SVP is not significant at the 5% level ($p=0.54>0.05$). In other words, SVP does not outperform LSVP for static targets. This conclusion – that the moving distractions did not distract more with LSVP than with SVP is compatible with the findings of Holcombe (2010) and Abrams (2003) that, unless the movement of a subject is 'odd' or unusual in some way, it will not attract a participant's attraction more than if it were static.

Subjective questionnaire

Responses to questionnaires revealed interesting facts about the opinions of participants concerning the three viewing modes. Responses to the questions "How easy was it to perform the task?" and "How confident are you that you found all the targets?" (Figure 8) provide a measure of the confidence and enjoyment with which the participant performed the test. Note that SMP shows a lower level of confidence than the other modes.

Figure 8 (Left): measures of participant confidence, (right): Measures of participant enjoyment





Experiment 2 (low distractor density)

As mentioned earlier, experience strongly suggested that a factor significantly affecting the relative advantages of the viewing modes SMP, SVP and LSVP was the density of distractors. A second experiment was therefore carried out using a distractor density significantly lower than used in Experiment 1. Figure 9 shows a view of the environment employed in Experiment 2.

Figure 9: A view of the low-distractor environment

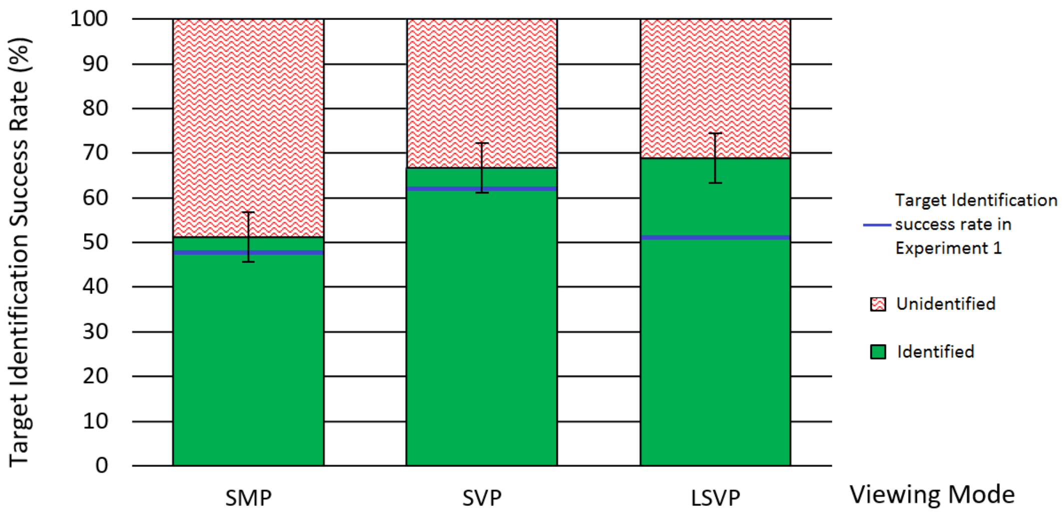
Other than the lower density of distractors, the design of Experiment 2 is identical to that of Experiment 1 but with two differences. First, with the help of 10 participants the difficulty of the targets was modified to bring the overall identification success rate closer to 50% in order to be comparable with Experiment 1. Second, most of the 48 participants were IT specialists and engineers in contrast to members of the public who took part in Experiment 1.

The following hypothesis was proposed:

H4: With the lower distraction density, LSVP will, overall, lead to higher target identification rate than SMP or SVP.

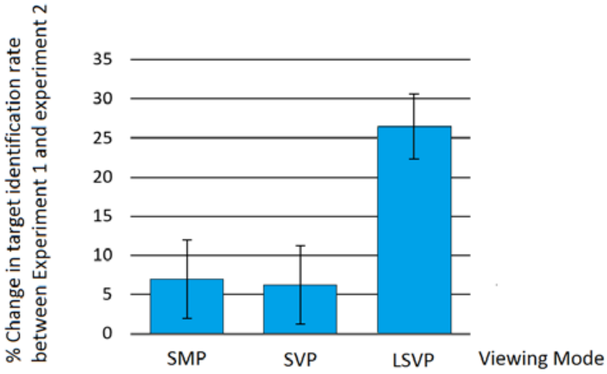
Figure 10 shows the results for the three viewing modes, using the same mixture of static and moving targets, as well as the same targets, as in Experiment 1. With SMP, participants correctly identified 51.3% (SD=18.1%, N=96) of the targets; with SVP the identification level was 67.5% (SD=17.9% N=96); and with LSVP 68.1% (SD=17.2%, N=96) of targets were correctly identified. Statistical analysis using the T-test indicates that the hypothesis that LSVP is superior to SMP is significant at the 5% level ($p=0.0002<0.05$). It also establishes that LSVP is not significantly superior to SVP at the 5% confidence level ($p=0.51>0.05$). Thus, LSVP does not lead to a significantly higher success rate than SVP, though it represents a significant improvement over SMP when compared with the higher distraction level of Experiment 1 as shown by the blue bars in Figure 10.

Figure 10: Overall target identification rate for the three viewing modes, observed in Experiment 2 using a lower distraction density.



The question naturally arises as to the extent to which distraction level influences target identification for the three viewing modes. The result of this investigation is shown in Figure 11. With SMP the percentage changes to the identification rates were 6.97%, for SVP it was 6.5% and for LSVP it was 26,4%. Statistical analysis indicates that LSVP provides a significantly greater improvement in target recognition in comparison with both SMP and SVP when compared with the higher distraction level. Hypothesis **H4** is therefore **confirmed**.

Figure 11: Percentage change in target identification rate between Experiment 1 (high distractor density) and Experiment 2 (low distractor density).



The result shown in Figure 11 is significant in that it points to the sensitivity of target identification to distractor density, and suggests that, with even lower distraction levels, the LSVP mode may be even more superior to SMP and SVP. If this were true then, for example, LSVP has the potential to perform well in scenarios – such as Wilderness Search and Rescue – where the distractors may be at a much lower level and where it is essential to identify a target that is moving in some way (e.g., waving). The rate of false positives for each viewing mode was recorded. Generally the rates were marginally higher than for the high distractor situation, with the rates for SMP being noticeably lower than for SVP and LRSVP.

Moving targets

In view of the interest in LSVP the identification rate for moving targets alone was observed for the three viewing modes (Figure 12). Statistical analysis established that LSVP performs better than SVP and SMP for (moving) targets 4 and 6, but worse than SVP for moving target 3. Examination of the nature of target 3 subsequently identified a flaw associated with the period for which it was visible, so the statistical analysis was repeated with target 3 removed. The result showed that, with this removal, LSVP (average 87%) was significantly better than SVP (58.7%) and SMP (54%).

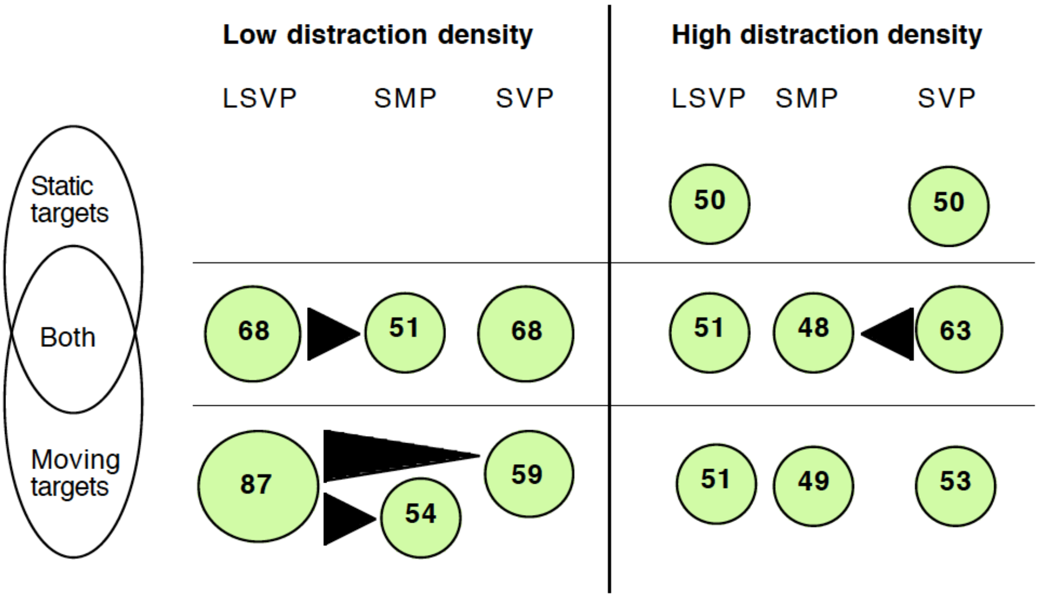
Figure 12 Target identification success rate for moving target in the low distraction density environment of Experiment 2.

Discussion

Figure 13 summarizes the findings obtained from statistical analysis of the experimental results regarding target identification success. Significant confidence levels (5%) are indicated by the black arrows. For high levels of distraction density (on the right) three groups of results are shown according to whether the targets were solely static or moving, or were a mixture of those two types.. Because the results were obtained for representative samples of each of the three viewing modes (LSVP, SVP and SMP) no boundaries between the three groups are shown. Similarly, distraction levels have not been quantified. Examination of the results for high distraction levels suggests that SVP might usefully be chosen as a default whatever types of target are involved.

On the left of Figure 13 we see a different picture. With only static targets there is no reason to favour any one of the three viewing modes, whereas with moving modes alone there is strong evidence that LSVP is superior to both SVP and SMP. For the equal mixture of moving and static targets LSVP and SVP are equally effective under the conditions tested.

Figure 13 Summary of the statistical analysis of the results obtained in Experiment 1 (high distraction density) and Experiment 2 (low distraction density). Black arrows indicate statistical significance at the 5% level. Circle diameter is proportional to average target identification, the latter also being indicated numerically. The result for moving targets at low distraction density ignores one exceptionally challenging target.



In the course of the investigation reported above, various limitations to the generalisation of the results were identified. They include the design of targets – especially the duration of their visibility – and the specific features of pedestrian appearance and animation.

Future research

The existing literature on Search and Rescue identifies a wide range of fundamental questions requiring research. The fruitful lines for future research noted below focus on the topic of this paper.

Recording of a spotter's eye gaze movement would yield data whose study could beneficially influence the future development of viewing modes. Useful information about the reason for false positives might also emerge. Similarly, the form of interaction used to identify targets (e.g., touch, gaze, mouse) would benefit from experimental study. Also, the potential influence of gaze on the temporary slowing down or magnification of tentatively identified targets, as investigated by Mardell et al (2011) could benefit from study.

Freedom of choice of the camera's spectral range must also be investigated (e.g., infrared)

Automatic target recognition (both true and false) and its appropriate presentation to the spotter is an additional route for future research.

Conclusion

Two principal conclusions can be drawn from the experimental results.

First, and with two significant exceptions, target identification varies little between viewing modes. One exception is the identified superiority of SVP over SMP at high distraction density and which tends to confirm earlier work by Mardell et al (2013). The other is the newly discovered superiority of LSVP at low distraction densities. There would appear, therefore, to be potential for the use of LSVP when targets may be moving, particularly at low distraction levels, and its further investigation is encouraged by the observed sensitivity of target identification to distraction density (Figure 11)

Second, the use of game development software appears to support the simulation of a search scenario sufficiently realistically to support useful experimental investigation.

Abbreviations

SAR	Search And Rescue
SVP	Static Visual Presentation
LSVP	Live Static Video Presentation
SMP	Standard Moving Presentation
SD	Standard Deviation
N	Number of trials for statistical analysis

References

Abrams, R. A. (2003). Motion onset captures attention. Psychol Sci., (5), pp. 427-432.

Adams, J.A., Humphrey, C.M., Goodrich, M.A., Cooper, J.L., Morse, B.S., Engh, C. and Rasmussen, N. (2009) Cognitive Task Analysis for Developing Unmanned Aerial Vehicle Wilderness Search Support, Journal of Cognitive Engineering and Decision Making, 3, 1, pp.1-26.

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.

Doherty P., Rudol P. (2007) A UAV Search and Rescue Scenario with Human Body Detection and Geolocalization. In: Orgun M.A., Thornton J. (eds) AI 2007: Advances in Artificial Intelligence. AI 2007. Lecture Notes in Computer Science, vol 4830. Springer, Berlin, Heidelberg.

Drew, T., Vo, M.L.H. and Wolfe, J.M. (2013) The invisible gorilla trikes again: sustained inattentional blindness in expert observers, Journal Psychological Science, 24, 9, pp.1848-1853.

Goodrich, M.A., Morse, B.S., Gerhardt, D., Cooper, J.L., Quigley, M., Adams, J. and Humphrey, C. (2008) Supporting Wilderness Search and Rescue using a Camera-Equipped Mini UAV. Journal of Field Robotics, 25, 1, pp. 89-110.

Healey, C. and Enns, J. (2011) Attention and Visual Memory in Visualization and Computer Graphics, IEEE Trans. Visualization and Computer Graphics, 18, 7, pp. 1170-1188.

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

Mardell, J., Witkowski, M. and Spence, R. (2013) A comparison of image inspection modes for a visual search and rescue task, Behaviour & Information Technology, pp. 905-918.

Mardell, J., Witkowski, M. and Spence, R. (2011) Gaze-contingent enhancements for a visual search and rescue task, APGV'11 Proc. ACM SIGRAPH Symposium on Applied Perception in Graphics and Visualization, page 109.

Simerson, R. (2004) Civil Air Patrol Mission Aircrew Scanner Course.
Stager, P. a. (1975). Eye-Movements and Related Performance in SAR Visual. Defence and Civil Institute of Environmental Medicine.

Waharte, S., Trigoni, N. and Julier, S.J. (2009) Coordinated Search with a Swarm of UAVs, Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks Workshops.

Waharte, S. and Trigoni, N. (2010) Supporting Search and Rescue operations with UAVs, International Symposium on Robots and Security, (Book).

Ware, C. (2012) Information Visualization: Perception for Design, Burlington, MA, Morgan Kaufmann
Wolfe, J.M. and Pashler, H. (1998) Visual Search, Attention, 1, pp. 13-73.

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.

Wolfe, J.M. (2004) Guidance of Visual Search by Preattentive Information, Chapter 17 of Neurobiology of Attention, 2005, pp.101-104.

Wolfe, J.M. and Gray, W. (2007) Guided Search 4.0, Integrated Models of Cognitive Systems, pp. 99-119.

Wolfe, J.M., Aude, O., Horowitz, T.S., Butcher, S.J. and Bompas, A. (2002) Segmentation of objects from backgrounds in visual search, Vision Research 42, 28, pp. 2985-3004.