Movers, Shakers, and Those Who Stand Still: visual attention-grabbing techniques in robot teleoperation

Daniel J. Rea, Stela H. Seo, Neil Bruce, James E. Young
University of Manitoba
{daniel.rea, stela.seo, bruce, young}@cs.umanitoba.ca

ABSTRACT

We designed and evaluated a series of teleoperation interface techniques that aim to draw operator attention while mitigating negative effects of interruption. Monitoring live teleoperation video feeds, for example to search for survivors in search and rescue, can be cognitively taxing, particularly for operators driving multiple robots or monitoring multiple cameras. To reduce workload, emerging computer vision techniques can automatically identify and indicate (cue) salient points of potential interest for the operator. However, it is not clear *how* to cue such points to a preoccupied operator – whether cues would be distracting and a hindrance to operators – and how the design of the cue may impact operator cognitive load, attention drawn, and primary task performance. In this paper, we detail our iterative design process for creating a range of visual attention-grabbing cues that are grounded in psychological literature on human attention, and two formal evaluations that measure attention-grabbing capability and impact on operator performance. Our results show that visually cueing on-screen points of interest does not distract operators, that operators perform poorly without the cues, and detail how particular cue design parameters impact operator cognitive load and task performance. Specifically, full-screen cues can lower cognitive load, but can increase response time; animated cues may improve accuracy, but increase cognitive load. Finally, from this design process we provide tested, and theoretically grounded cues for attention drawing in teleoperation.

CCS Concepts

• Human-centered computing ➝ Interaction design ➝ Interaction design process and methods ➝ User interface design

Keywords

Human-robot interaction; multi-robot teleoperation; attention;

1. INTRODUCTION

Improved robot teleoperation interfaces for applications including the military [38], industrial [42], or domestic tasks [27,29], remains an ongoing research challenge. To improve operator efficiency, a goal of teleoperation is to enable fewer people to control or monitor more robots, increasing the human-robot ratio [31,53], and getting more work done faster. Unfortunately, increasing the information given to operators, such as simultaneous video feeds from separate cameras or robots, results in higher cognitive load and operator error (e.g., [12,35,36,42]). As such, a primary goal of teleoperation usability research is to improve overall operator effectiveness: provide them with the tools and information they need to perform their tasks, without overloading them mentally. We follow this theme, exploring tools to increase performance on visual search tasks.

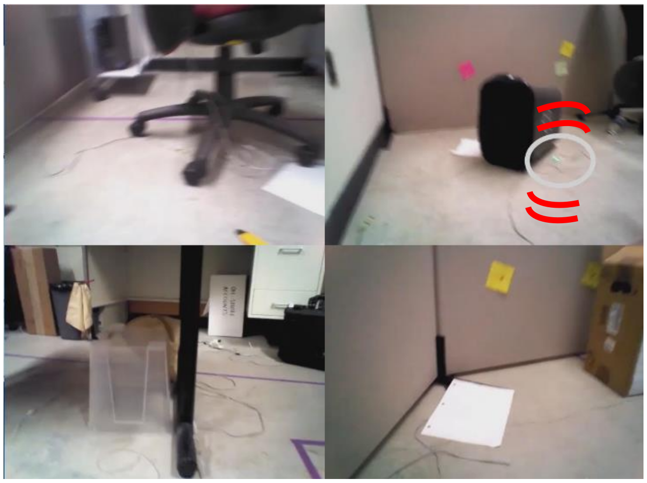

Figure 1. Visually cueing a point of interest on a tiled camera feed, e.g., by using a bouncing circle that draws attention to a point of interest (green light, red lines indicate circle movement, and are not shown in the interface).

Emerging computer vision techniques (e.g. [7,47]) can help operators in visual search tasks by automatically identifying potential points of interest, and indicating (*cueing*) the points for inspection. However, it is not yet clear how this information should be cued to an operator to effectively gain their attention, without distracting from the primary task. This balance is not obvious: cues cannot be too subtle, as while attention is focused (e.g., on a search task) people may not notice events outside their immediate focus [50]. Cues that are too intrusive can be annoying, frustrating, and distracting to the point of negatively impacting the primary search task [46].

Drawing from psychology literature on human attention, we conducted an iterative design exploration of using visual cues in a multi-robot search and rescue context. We iteratively designed, implemented, and evaluated cue variants based on cue proximity (at a target point or full-screen) and cue motion (moving or static), while measuring operator visual-search accuracy, response time, and cognitive load. Results indicate that moving cues can help operators find more lights than static cues. Further, cues located at a light, particularly when moving, can be the quickest for operators to assess, but also increase cognitive demand. Well-designed full-screen cues can achieve similar task effectiveness while simultaneously lowering the cognitive load required on operators. In addition to our study results and reflection on these parameters, we present a set of tested, iteratively designed cues that aide operators in visual search tasks without negatively impacting the impact on cognitive load.

2. RELATED WORK

Supporting teleoperation (aiming to increase performance and lower cognitive demands) by modifying how video feeds are presented is an active research area [12,17,27,38]. Much of this has

HRI '17, March 06-09, 2017, Vienna, Austria
© 2017 ACM. ISBN 978-1-4503-4336-7/17/03…$15.00
DOI: http://dx.doi.org/10.1145/2909824.3020246

revolved around camera location and viewpoint choice, for example, egocentric views enable an operator to see the world from the robot's perspective [12,26,27], environmental views provide robot-in-context information [17], while bird's-eye and third-person views provide this context calibrated around the robot as it moves [14,39,40]. When multiple feeds are displayed, research has explored layout options, such as displaying all screens in a tiled fashion to maximize available information [44], or using picture-in-picture techniques to prioritize screen real-estate toward more important views [24]. Feeds can be hidden until requested [15], with operators perhaps rotating through them [3] – reducing information load and saving screen space [15]. Rather than projecting views or representing camera placement, we aim to support teleoperation by providing task-specific help: drawing operator attention to points of interest while mitigating the negative effects of the interruption.

The video feed itself is commonly augmented with graphics to represent relevant information, such as sensor data [22,38], or task information as with ecological design [17,29,42], including notifications [9]. We expand on this work, by investigating how to notify users of potential points of interest within a video feed.

In computer vision, saliency detection refers to the problem of detecting image regions that are likely to be salient to a human viewer. This is a complex problem which considers physiology of vision as well as psychology. Saliency detection has been used successfully to shift [52] and predict [7,49] gaze, improve visibility in augmented reality [21], and model analysis of context [10]. It has also been used to modify video scenes, and to draw human attention toward objects (e.g. [52]); in this case, the changes were subtle and the goal was for impacting viewer tendency to look at areas, not direct attention to immediate concerns. For teleoperation, saliency detection has been applied only rarely, to inform interface design of sensor readouts [9], or to minimize transmission bandwidth [47]; here, low-saliency regions use lower resolution, effectively blurring them. We extend this work by investigating how visual cues can be designed to appropriately draw operator attention to such points identified through saliency detection, while aiming to lower operator distraction that may impact task performance.

Human visual attention, particularly for visual search, has a rich history in psychology. Studies frequently focus on static abstract images [10,23,25,48], or abstract videos with colors or shapes changing and moving against solid backgrounds [1,3,8,21–23]. Studies with natural scenes tend to use static images [10,49,50]. Some video work in natural scenes focuses on closed-circuit television monitors [3,11,20,43]. Robots, unlike CCTV cameras, move throughout their environment freely, presenting highly dynamic and noisy camera views from constantly shifting perspectives; further, increasingly dynamic environments tax users' attention resources [32,45]. As such, prior attention results must be specifically evaluated in the unique, high-demand teleoperation context.

Notification work for general desktop applications has focused on *when* to draw attention [4,18,19,30], which does not apply to our task, where operators must be immediately notified and respond in a short time (before a target leaves the screen). Work on *how* to draw attention is much more limited. Some has highlighted how interruptions can be distracting or annoying, and has investigated how to minimize these problems [52], while (recently) noting the lack of solutions to this problem [46]. Further, heavy use of interruptions can lead to users ignoring them [8]. These results, primarily from web and desktop applications, motivate the need for our research in the visually-intense and noisy teleoperation task, exploring attention-drawing cue design that balances being attention-grabbing while not being distracting or being ignored due to fatigue.

3. ATTENTION AND PERCEPTION

Human visual attention – how people choose what to focus their vision resources on – is a well-studied topic in biology, neurology, psychology, etc. Attention can be defined as an enhanced response to stimuli at an attended location and, as a result, reduced response to stimuli elsewhere [50]. Thus, we can expect people to have increased focus on some task elements (e.g., during searching, driving, reading), and, inversely, difficulty noticing things outside of their focus [37], even highly-salient points of interest – this is called inattentional blindness [28]. This is especially difficult during noisy dynamic tasks, such as teleoperation [41]. We aim to work within human patterns of attention to devise visual mechanisms to help gain people's attention and direct it to points of interest, with minimal overall hindrance or additional strain on cognitive resources.

One technique for focusing attention, called goal-based attention, cognitively directs attention to known criteria or stimuli, such as a known suspect on CCTV [20]. Goal-based attention is relevant to teleoperation as operators often have specific, if broadly defined, tasks that drive visual search and cognition such as "find and rescue all victims in a disaster." Complicated search goals (multiple criteria, complex shapes) and environments, such as disaster environments, reduce the effectiveness of goal-based attention [20,37,51]. Increasing the number of cameras will also reduce the effectiveness of goal-based attention due to the increased search area [43]. Goal-based attention quickly reaches limitations in complicated tasks and environments that may be present in tele-robotic search and rescue.

Alternatively, stimulus-driven attention draws a person's attention to salient stimuli, such as bright lights, motion, or high contrast graphics. Interfaces could have objects appearing [13], elements starting to move [1,2], or motion perpendicular to other motion in the visual field [13]; not all changes are similarly salient, for example, color shift, or motion types such as receding motion or movement parallel to other ongoing motions, have been found to be less effective at drawing attention [1,13]. Stimulus-driven design suffers less from fatigue in comparison to goal-based attention, and further suffers less from inattentional blindness, important for long-term attention (vigilance) [11]. As such, we design cues leveraging stimulus-based attention to mitigate some of the limitations incurred by the operator's goal-driven attention.

4. CUE DESIGN PROCESS

Our investigation into how cue design impacts teleoperator performance employed an iterative design process: we drew from perception and attention literature to inform design, devised a mock urban search and rescue task for evaluation, implemented our cues into the mock task, and conducted formal experiments to learn of the impact of our cue designs. Our results informed the design of new cues and conducted more experiments, for a total of one pilot (9 participants) and two formal studies (with 20 participants each).

4.1 Cue Evaluation Test Bed

We developed a test bed that engages participants in mock urban search and rescue, performing visual search on teleoperation feeds, enabling us to test the impact of our cues on visual search.

4.1.1 Task

Participants monitored a collage of four tiled video feeds from teleoperated robots exploring a mock-disaster environment (Figure 1), and were asked to search for stimuli that represented points of interest (e.g. potential victims, dangerous equipment). Participants tapped the screen near the stimuli to show they had identified it.

For our stimulus, we aimed for an abstract stimulus that would more readily generalize to a broad range of tele-operation tasks. As such, we avoided being domain-specific, and potentially confounding variables such as shape or pattern. Our design goal was for an abstract, generalizable stimulus which is unambiguous once found, yet still difficult to find. We chose green point lights as our target. Further, we aimed to increase visual search validity by using a visually noisy scene with realistic robot movement and video quality, building on existing fully abstract perception work by investigating attention in a more representative visual environment.

4.1.2 Teleoperation Videos

We pre-recorded our robot teleoperation videos for consistency across participants. As participants only monitored the feeds, and did not actually see the teleoperators, this is equivalent to live operation for our evaluation purposes.

We arranged a room to have furniture, electronics, and debris scattered around (Figure 1), and remotely controlled a NAO H25 robot over Wi-Fi traversing the space. The video was recorded from the robot's head camera (640x480 at 17 FPS).

We recorded five videos, each having a unique room and debris arrangement, while maintaining similar visual clutter, layout, lighting conditions, and robot movement properties (speed, frequent turns, minimal stopping). We compiled five different (but comparable in character) four-tile collages for a repeated-measures study design. We modified the video selection and position in the collage using incomplete Latin Squares to minimize learning effects. Each video and collage lasted six minutes and four seconds long.

4.1.3 Stimuli (light) Placement and Timing

We placed several centrally-controlled green LED lights throughout the mock environment to serve as our stimuli. Light timing and placement posed several challenges. First, only one light at a time should be illuminated in the entire collage, to avoid confusion over which stimulus a participant noticed. As such, lights could not simply be left on, and had to be triggered as needed. Second, lights should not turn on or off in-scene, as this change itself is a confounding stimulus [13], and should change off-camera. Third, there should be a consistent minimum delay between the stimuli (but not fixed, to avoid predictability), to avoid confusion over which light a participant responds to (we used one second). Finally, light occurrence between the videos in a collage should be evenly balanced.

The coordination of lights turning on and off within a video, and between videos, was non-trivial, particularly given how videos would be combined into various collage configurations. We employed a master schedule that dictated light timing, and made minor imperceptible changes to video speed to ensure all constraints were met. Each video was over six minutes and four seconds long and had exactly 12 light stimuli. Thus, each collage had exactly 48 light stimuli, which showed up on average every 8 seconds. As each video had a unique stimuli timing, the relative timing between the stimuli changed in each collage due to our Latin Square balancing.

4.1.4 Cue Integration into Video

All visual cues were created using post-processing in Adobe Premiere and After Effects. For each experiment, a full set of collage videos were made for each cue (each collage had a version with one cue type applied) to allow for within-participant counterbalancing.

Rather than simply attaching cues to all lights in a video collage, for improved ecological validity we also included false-positive cues (cue without stimuli), false-negative cues (stimuli but no cue), and cue misses (a stimulus, but cue at an incorrect location). These not only simulate the realities of imperfect saliency-detection systems [5], but were designed to imbue a sense of diligence in participants, as they could not completely trust the cueing system. Further, introducing cueing errors into a system, while realistic, can have overall detrimental effects on performance: operators can lose trust in unreliable cues and overcompensate with increased attention, introducing additional error [2,13,33,34], potentially more than an un-cued case. As such, these standard errors must be part of a test bed for comparing cues to an un-cued base case.

In our case, each collage contained 52 cue instances: 40 correct true-positive cues (77%), 4 false-positive cues, 4 false-negatives (no cue), and an additional 4 cue misses, for a total of 23% error cases. False cue rates were based on prior attention work [33].

4.1.5 Instruments

Participant taps (indicating they saw a light) produced a short beep to indicate it was registered, and were recorded and automatically processed for response time and accuracy. Accuracy was further broken down into correct identification of a light, tapping with no light or cue, and tapping the cue and not the light in the mis-cue case (cue in wrong location). A tap was correct if it occurred in the correct feed within 2 seconds of the light disappearing (a generous upper limit based on an expected maximum .5s reaction time [34]).

After each task (i.e., with one cue), participants filled out a short questionnaire to measure subjective cognitive load (NASA TLX [16]), and custom 20-point Likert-like items (mimicking the appearance of the TLX scales) for nausea, trust in the interface, enjoyability, and self-perception of speed at the task.

At the end of the experiment, participants answered a free-form short answer section on pros and cons of each cue, as well as any comments on any motion sickness, or other comments.

Participants sat in front of a Microsoft Surface 2 tablet, with the video collage displayed in full-screen and at max brightness, with the minimum tilt setting. Participants were not allowed to pick up the tablet or change the tilt. The desk, chair, and tablet displaying the collages that participants used were placed at fixed initial positions, though participants could adjust the chair to be comfortable.

4.1.6 Procedure

Before beginning the experiment, participants are briefed on the task before reading and signing an informed consent form. They were then given a 30-second practice collage to watch, and shown how to indicate where in the collage a light appeared (by tapping). They were told that the videos were pre-recorded using real robot, and were informed of, with examples, of how the cueing system sometimes made mistakes (false cues).

Before starting the tasks, participants are shown example collages containing all visual cues in the order they would appear in the experiment. Participants were told (and reminded before each task) to act as quickly as possible as time was being recorded.

The experiments used within-participants design, with participants completing the task with all cue designs; cue orders and cue-collage mapping were counterbalanced using incomplete Latin Squares.

Before each task we displayed the cue to refresh the participant's memory, and the task started when the participant touched the screen). Between tasks, before moving on to a new que, there was a mandatory three-minute break to mitigate the impact of fatigue; during this time, participants filled out the post-task questionnaire.

After all tasks were complete, participants completed the post-experiment questionnaire, and were debriefed on the experiment.

4.2 Cue Design

Our cue-design methodology was based on our background exploration in human attention literature, as summarized in Section 3. As motion is highly effective at drawing attention [1,2], it is a strong candidate for cue design. However, motion can be distracting [46], and may have a negative impact on primary task performance. Further, in our search and rescue application, the visual field is already noisy: constantly changing as the robot navigates; we need to investigate if motion cueing is still effective in this scenario, or, if the combined motion of the cue and robot becomes even more distracting. We investigate cue motion as a design variable: cues that move (*moving* cues), and cues that do not move (*static* cues). As the light is always moving in the visual field of the robots, we defined static cues to be fixed relative to the moving light.

On-screen cue location is important as it impacts the cue visibility: an operator may be focusing elsewhere when a light appears. Cues located near a light encode the location of the light and thus reduces the search space once noticed [43]. Therefore, we may expect these cues to elicit fast response times, as once a cue is seen, an operator does not need to search for the light. However, due to inattentional blindness, operators may not notice even highly salient cues outside their current attention [41], and so we examine full-screen cues (visible everywhere at once). These should be easy to notice, no matter where an operator is focusing, which indicates to the operator that a target is currently on screen. Therefore, we investigate the cue proximity as our second design variable: cues at the light (*at-light* cues) and cues over the entire visual field (*full-screen* cues).

We use these two design variables, cue motion and cue proximity, to drive our cue design as well as evaluation.

5. INITIAL CUE DESIGN AND PILOT

We conducted an initial pilot study as a broad exploration into cue design for supporting teleoperation visual search, using our two design variables: cue motion, and cue proximity. In the pilot, our full protocol was not followed: we only measured accuracy, and we used an earlier and rougher video collage that was longer, had less rigorous light spacing, and had all cue types intermixed.

5.1 Initial Cue Design

We designed and implemented an initial set of nine cues based on our perception literature exploration and our two design variables.

Our initial at-light (cue proximity variable) cues were *red circle* and *grey circle*, simple outlines, and *exposure*, a disc of increased exposure, over the stimulus. These were chosen to explore the impact of visual contrast, a known factor in salience [11,20].

For investigating movement, we animated the grey circle to bounce one cue radius either left to right (*vertical* cue) or top to bottom (*horizontal* cue). Motion direction, either parallel to or orthogonal to visual flow, can impact salience [1,13]; given that our robots turn often but do not look up or down frequently (except when they fall), *horizontal* is parallel and *vertical* is orthogonal.

For the full-screen component of cue proximity, we aimed to impact the whole visual field, to be difficult to ignore, while simultaneously trying to indicate where the light is. We tried blurring (*blur* cue) or darkening (exposure reduction) the entire screen except for a disc around the light. Both changes in clarity and exposure have been shown to be salient [6,52]. These cues cannot trivially be made dynamic, as simply animating them would not be effective as both the blur and darken effects would not show change as they move.

For our full-screen, moving cue, we drew from video-game design and implemented a common targeting animation: *target* was a circle approximately the size of the screen that appeared and rapidly shrank towards the target. While the shrinking motion should attract attention, particularly as a shrinking circle has orthogonal motion to all visual flow directions [1,13], a risk is that it may appear as a receding motion, which has been shown to be less salient [1]. For all moving cues, the animation lasted for 1 second for consistency across cues; horizontal and vertical bounce cues and target cues all became static grey circles until the light left the screen.

5.2 Pilot Study

Our primary focus of the pilot study was to direct our exploration for more formal study by evaluating our test bed, testing our design variables, and obtaining an initial sense of our cue design successes and failures. As such, we ran our pilot with all eight initial cues (red circle, grey circle, exposure, horizontal bounce, vertical bounce, blur, darken, and target) and compared their impact on how many lights participants correctly identified (accuracy). While we compare data across all cues, of particular interest to us was which cues performed best in each design configurations: at-light static, at-light moving, full-screen static, and full-screen moving.

In addition, we added a case with no cueing (just the light stimuli): this was to measure the overall impact of cueing (e.g., can possibly make performance worse), as well as to test the task itself, to ensure that it was sufficiently difficult to benefit from cues.

5.2.1 Results

We conducted our pilot with nine participants recruited from our general university population. A one-way repeated measures ANOVA showed an effect of cue type on accuracy ($F_{8,64}$=5.71, η^2=0.42, p<.001, Figure 2). Post-hoc comparisons (with Bonferroni correction) comparing all cues to each other revealed the best performing cue for each design parameter set. While Target was the only full-screen moving cue, it was validated by performing statistically better than four other cues. If there was not a clear winner, we simply picked the cue with the highest mean.

In the no-cue case, operators found on average 66% of lights (std. dev 17%). This was comparable to some of the worse cues (exposure, horizontal bounce). This indicates that our test bed and visual-search task are sufficiently difficult, where the addition of cues can potentially improve on the success rate. However, some cues seem to perform at least as badly as having no cue at all (Figure 2).

6. FIRST DESIGN ITERATION

The results of our pilot study gave us initial representative candidates for each design parameter combination: grey circle performed best as our at-light, static cue, the full-screen dark effect as our full-screen, static cue, the vertical bouncing circle as our at-light moving cue, and the video-game inspired targeting as our full-screen

Table 1. Representative cues for testing our cue proximity and movement design parameters.

		cue proximity	
		at-light	full-screen
cue movement	static	circle	dark
	moving	bounce	target

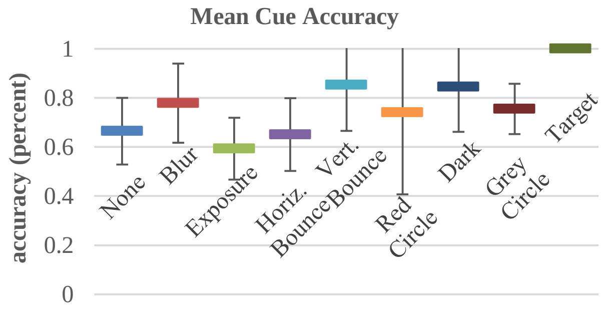

Figure 2. Mean accuracy of our initial cues. Error bars show 95% confidence intervals.

moving cue. We refer to these as *circle*, *dark*, *bounce*, and *target*, respectfully, as summarized in Table 1 and shown in Figure 3.

We employed the full test-bed protocol as a within-participants experiment: each participant completed the task with each of the four interfaces. We counterbalanced cue and collage order.

The purpose of this iteration was to more formally and rigorously test our design variables, cue proximity and movement, using our candidate representative cues developed through the exploratory pilot study. We again included the no-cue case to more rigorously test the overall impact of cueing in comparison to the un-cued base case (e.g., cueing may possibly hinder performance).

6.1 Results

We recruited 20 participants (8 female) from the local university population. The mode age (collected in ranges) was 26-30, at 35%.

Repeated-measures ANOVA comparing all cue against the no-cue case) showed an effect of cue type on response time (Figure 4b, $F_{2.8,52.3}$=41.9, η^2=.69, p<.001, Greenhouse-Geisser correction), accuracy (Figure 4c, $F_{2.0,38.3}$=30.8, η^2=.62, p<.001, Greenhouse-Geisser correction), and cognitive load (Figure 4a, $F_{2.2,41.8}$=6.5, η^2=.26, *p*=.003, Greenhouse-Geisser correction). Planned contrasts against no cue showed all others to be more accurate and to have lower cognitive load (*p*<.002), while circle, bounce, and dark had faster response time; no response-time difference was found against target (*p*<.01). While Figure 4 shows overall means and confidence intervals, the within-participants statistics uses relational scores.

We performed 2-way repeated-measures ANOVAs (cue proximity X motion) on operator accuracy, response time, and cognitive load (Figure 5). Bonferroni-corrected post-hoc tests were performed (with main effects) to investigate the effect for each of the levels.

We found a main effect of cue proximity on operator response time: at-light was faster than full-screen (Figure 5b, $F_{1,19}$=107.3, η^2=.85, *p*<.001, 95% CI [-232ms, -154ms]). Post-hoc tests revealed that circle was faster than dark (*p*=.021, 95% CI [-153ms, -14ms]), and bounce was faster than target (*p*<.001, 95% CI [-353ms, 252ms]).

We also found a main effect of cue movement on response time revealing that static was faster than moving, (Figure 5b, $F_{1,19}$=4.9, η^2=.20, *p*=.04, 95% CI [-113ms, -3ms]). Post-hoc tests revealed that dark was faster than target, p < .002, 95% CI [-235ms, -98ms]; static circle versus moving bounce was n.s. There was an interaction effect between the two parameters ($F_{1,19}$=24.3, η^2= .56, *p*<.001).

We found a main effect of cue proximity on operator accuracy, revealing that operators found more lights with at-light cues than with full-screen cues (Figure 5c, $F_{1,19}$=4.4, η^2=.19, *p*<.05, 95% CI [0 lights, 1.15 lights]). Post-hoc tests showed circle to have better accuracy than dark (*p*=.42, 95% CI, CI [-.23, -2.8 lights found]).

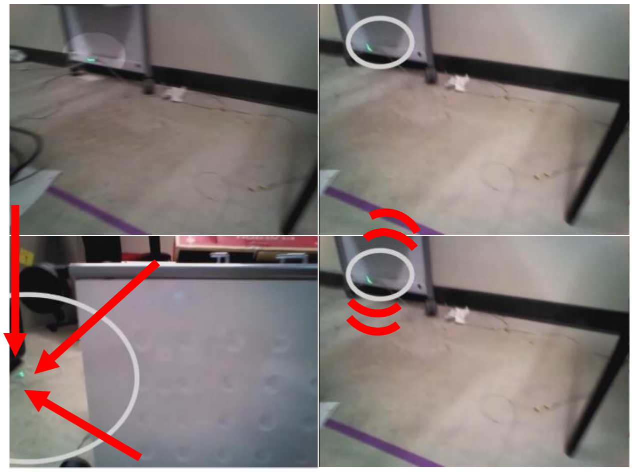

Figure 3. Our four interfaces for the first design iteration: dark (top-left), circle (top-right), target (bottom-left), bounce (bottom-right). Red markup indicates motion and are not shown in the interface.

We also found a main effect of cue motion on operator accuracy: operators found more lights with moving cues ($F_{1,19}$=6.5, η^2=.26, *p*<.05, 95% CI [-1.773, -.177 lights]). Post-hoc tests were all n.s.

We found a main effect of cue proximity on cognitive load, revealing that participants rated full-screen cues as demanding lower cognitive load (Figure 5a, *p*<.01, $F_{1,19}$=8.4, η^2= .31, 95% CI [-9.087, -1.463] NASA TLX points). Tests for main effect for cue movement, and interaction effects were all non-significant.

In summary, the 2x2 ANOVAs indicated that participants were faster and found more lights with at-light cues than full-screen cues, although the full-screen cues demanded lower cognitive load. Participants further were faster with static cues than moving cues, although they found more lights with moving cues.

No effects were found on nausea, trust in the interface, cue enjoyability, preference, or self-perception of task speed. Further, data on miscues and misclicks (when no light was present) were all n.s.

6.2 Analysis of Participant Feedback

To gain insight into strengths, weaknesses, and differences between cues, we performed open coding on short-answer feedback.

Motion cues (bounce and target) were seen by some as attention-grabbing (positive, 16 participants for bounce, two for target), while distracting by others (11 for bounce, seven for target), describing them as being "tiresome," and "stressful to eye," (bounce) or "breaking visual concentration," and "very distracting" (target). Some participants simultaneously praised the attention-grabbing

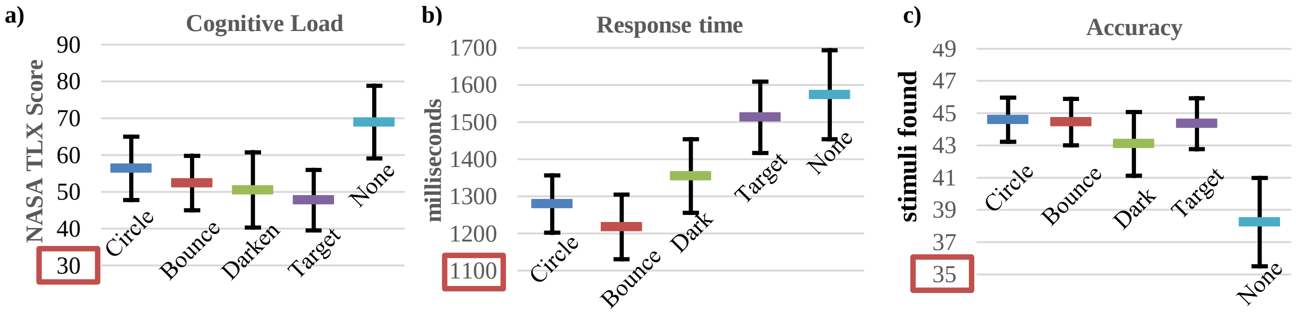

Figure 4. Results of planned contrasts, error bars are 95% confidence intervals. a) Cognitive Load Sum (range [5,120]): all cues performed better than no cue (*p*< .002). b) Response time: Only target was not better than no cues (*p*< .001). c) Mean Accuracy (range [0,52]): all cues performed better than no cues (*p*< .001).

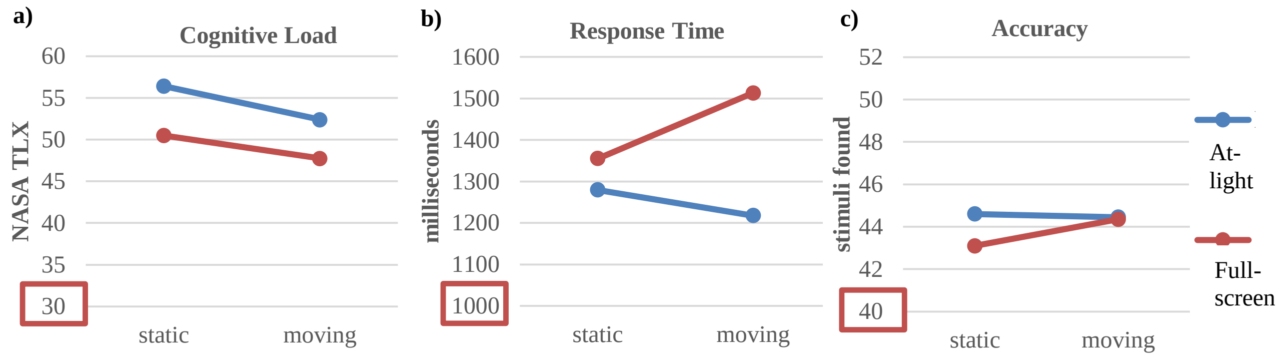

Figure 5. 2x2 (cue proximity X movement) results. a) Cognitive load (range [5, 120]): at-light cues had higher cognitive loads (*p*=.009). b) Response time (in milliseconds): at-light cues were faster (*p*< .001), and moving cues were faster (*p*=.04). An interaction effect is clear (*p*<.001). c) Accuracy (range [0,52]): at-light (*p*=.05) and moving (*p*= .019) had a higher accuracy.

properties of Bounce while commenting that it was too distracting. Static cues had fewer comments on distraction: four participants made comments such as the dark cue being "somewhat distracting," but seven noted that it was "minimally distracting." 11 participants noted that circle was "not easy to locate," and "not very distracting."

Participants commented that full-screen cues positively affected their comfort. Eight participants mentioned that the dark cue was "very relaxing," induced "less dizziness," and was "easy to visualize," and target cue had five comments about reduced stress "it highlights and easy to see. Less effort." Circle had six comments note it was "calm with a smooth motion" and that "appearing without a sense of movement diminished the urgency." No similar comments were given for bounce, but 11 participants complained it was tiring: "made me feel dizzy," "have to cautiously monitor four screens. Stressful" and, "heightened the sense of urgency."

Unique to the target cue, participants complained about its speed: five participants mentioned frustration: it was "too slow to capture the light," and "took focus off other screens for extended periods." No other cue had comments about speed.

Some participants noted the light-position information encoded in the full-screen cues. 10 mentioned that the target cue "helps you to target your focus on one [robot's video]" and that it "can tell me directly where the green light is when it's shrinking." While four people mentioned that with the dark cue the "increase light-to-background contrast made it easier to detect," eight participants conversely mentioned that it only "darkens the whole screen to let you know something has been captured" and required participants to "waste time by locating the image that is less darkened." There were no such comments found for the at-light cues.

Related to comfort, 14 participants complained about some level of nausea from the study. This was not linked to a specific interface, but was attributed by the participants to the simultaneous and often not coordinated movements of the robots.

6.3 Discussion

As with the pilot, our results confirm the validity of our test bed, as well as our cueing technique: all cues increased accuracy and lowered cognitive load, and all but the target cue increased response time (though Target was not found to be worse), when compared with no cue. Therefore, at the least, cueing may be useful to help participants in urban search and rescue tasks. Further, the benefits of our cues, for the results we measured, offset any detrimental effects that may be introduced, such as too much distraction [46].

The analysis of our two design parameters detailed important tradeoffs between design choices. Full-screen cue proximity appeared to demand lower cognitive load than at-light cues. It is not entirely clear why this may be, but participant feedback indicates that some found full screen cues more comfortable and readable, with target specifically being helpful for directing attention to a light. As well, bounce's motion (quick bounces up and down) is different from target's (smooth shrinking) which may also be a factor. Further, full-screen cues may reduce stress of potentially missing a cue, as they are much more difficult to miss.

At-light cue proximity resulted in operators finding more lights, and more quickly, than full-screen cues, despite the increased cognitive load. This can be explained in part by how at-light cues effectively make the stimulus larger. Further, at-light cues immediately indicate where the light is (once the cue is noticed). This can be contrasted with dark where participants complained they had to take time to search for it, or target, where they complained of the slow speed of our moving full-screen target cue which took an entire second to home in on the light.

Note that moving cues had slower response time. This is supported by our planned contrasts on response time: all interfaces except for Target were faster than None. Further, Figure 4b and Figure 5b suggests Target is the driving force behind the response times, performing worse than other cues, except for the no cue case.

For the cue motion design parameter, participants found a few more lights with moving cues in general, and overall were finding the lights faster than with static cues, although the specific results were mixed (the static dark was faster than the moving target). Participant feedback indicated that the moving cues were more salient, which explains the improved accuracy. The attitudes were again mixed, however, with some framing this positively as attention-grabbing or negatively as distracting, although the negative component was not reflected in performance or cognitive load scores.

Static cues were in general the poorest performers, with higher cognitive demands, and slower response times, and some participants noting that they were not easy to locate. This implies that Dark was specifically worse in Accuracy. Participants commented that they could easily see the Dark cue, but had to quickly look for the light itself: there was no position information encoded in the cue. This can be compared to Target which directly led the viewer to the light by the end of the animation.

6.3.1 First Iteration Summary

While the circle cue was better than no cue, and helped operators find the most lights, it appears to have one of the highest cognitive load demands, and did not perform well in response time (except or being faster than target) or accuracy. The dark cue similarly only performed well for speed in comparison to target, and had the lowest accuracy. Bounce was one of the strongest performers, with one of the fastest response times and best accuracies, but had complaints about distraction and fatigue, with improvements to be made in cognitive load. Target, while having speed issues, had one of the stronger cognitive load scores and was comparable on accuracy.

One important component of our analysis was the interaction effect found on response time (Figure 5b), which appears to be driven by the slow target cue. Target's poor speed was also reflected in it being the only cue that did not perform faster than no cue. Thus, we need to be careful about interpreting the response time main effects, as the specific target cue may need design improvements.

7. SECOND DESIGN ITERATION

We draw on our study results to develop a new set of cues to be tested. Some of these are iterations on our previous designs, while others are new designs based on our results from the first iteration. Our main goal in the second iteration was to develop hybrid cues with both static and moving elements, as well as full-screen and at-light elements to see if combining our cue design parameters could improve operator performance.

While the bounce cue was a strong performer, the weak point was the fatigue and cognitive load. We iterate on bounce by adding a full-screen element to try and mitigate these issues, to embed the comfort and cognitive load gains associated with our other full-screen cues. Specifically, we add a border to the video feeds for the duration of the cue, without changing the bounce itself (Figure 6). We hypothesize that this *framed bounce* will have improved cognitive load scores over the previous bounce, without negatively impacting response times or accuracy.

Participants commented on the benefits of full-screen cues encoding the location of the stimulus as well as providing an alert. We designed a new full-screen cue that statically encodes the light position; we hypothesize that avoiding moving elements can reduce frustration (and help with cognitive load), while maintaining the accuracy and cognitive load benefits of the full-screen design. Specifically, we used a greyscale radial gradient (linear, dark at edges, light at center) centered over a light (Figure 6). We anticipate that eyes can quickly follow the gradient from anywhere on screen toward the light as if looking through a tunnel. This encodes location similar to the target cue, but without the time constraint. We hypothesize that this *tunnel* cue will improve the cognitive load and accuracy over bounce, while achieving similar with response time.

As a secondary agenda, we directly iterate on our target cue. We believe that the animation that target uses to shrink toward the stimulus could be much faster, and still maintain its positive characteristics (low cognitive load and strong accuracy). As such, we developed a *fast target* variant which animated three times faster (0.33s instead of 1s). We hypothesize that this improved target cue will maintain the accuracy and cognitive load of target (not do worse), while improving on the response time.

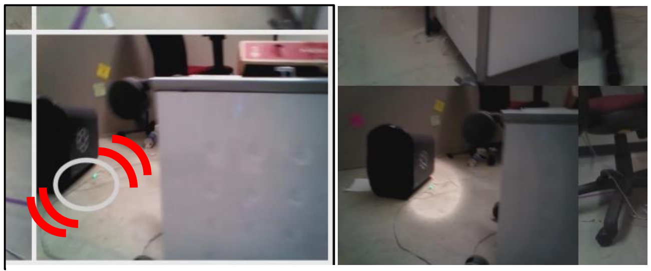

Figure 6. Our two new cues (from the left) framed bounce, and tunnel. Red lines indicate the animation

We conducted our full protocol using five cues: framed bounce, fast target, tunnel, regular target (to compare against fast target), and regular bounce (to compare against framed bounce, and tunnel). We keep target and bounce in the procedure to re-test for consistency and improved comparability with within-participants.

7.1 Results and Discussion

We recruited 20 people from around the university campus. Participant ages were collected in ranges; the mode was the 18-20 range, at 55%. We analyzed our data using t-tests given our targeted hypotheses, and did not use more exploratory methods. We summarize these results in Figure 7.

When comparing framed bounce to bounce, we found a trend for framed bounce to improve cognitive load (*t*=1.6, *p*=.064, 95% CI [-1.582, 11.682], one tailed). While no difference was detected between response times (*t*=-1.7,*p*=.11,[-96ms, 10ms]), framed bounce had better accuracy (*t*=-2.5, *p*=.021, 95% CI [-2.290, -.210 stimuli]).

Comparing tunnel to bounce, we found a trend for tunnel to improve cognitive load (*t*=1.5, *p*=.08, 95% CI [-1.771, 9.971], one tailed), with no difference found on accuracy (*t*=1.0, p=.16, 95% CI [-.725, 2.125 stimuli], one tailed). We found tunnel to be slower than bounce (*t*=-3.6, *p*=.001, 95% CI [-162ms, -43ms]).

We found that participants had a faster response time with our fast target, than regular target (*t*=7.6, *p*<.001, 95% CI [205ms 361ms], one tailed). We did not find difference for accuracy (*t*=0, *p*=1.0) or cognitive load (*t*=-0.9, *p*=.19, 95% CI [-8.503, 3.303]).

In this study, we successfully demonstrated how hybrid cues can be developed to integrate benefits of cues throughout our design space. Our full-screen plus at-light framed bounce cue had better accuracy

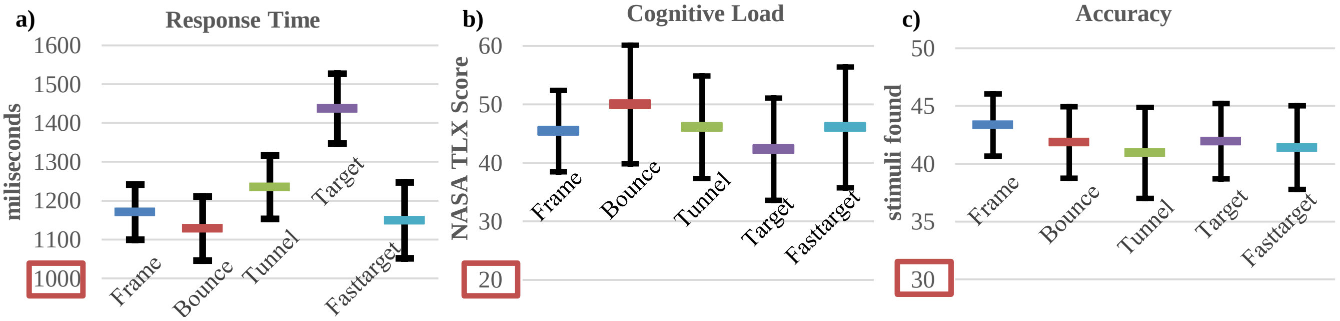

Figure 7. T-test results, error bars are 95% confidence intervals. a) Response time: Fast target was faster than target, bounce was faster than tunnel (*p*< .001). b) Cognitive Load (range [5,120]): there were trends for framed bounce (*p*=.064) and tunnel (*p*=.08) to incur less load than bounce. c) Mean Accuracy (range [0,48]): framed bounce helped more than bounce (*p*< .021).

than the regular bounce, comparable response times, and potentially improved cognitive load (a trend, requiring further study). At the same time, the failure of the tunnel cue, which may slightly improve cognitive load but harms response time, highlights the non-trivial nature of designing effective cues. It is likely our position encoding failed somehow, as it is a core concept in target, which performs well. Finally, we have demonstrated how our target cue can be improved simply by making it faster, negating many of the problems encountered with this cue in earlier study though our data points to a potential small effect of increased cognitive load.

8. Discussion

Looking across both studies (Table 2), motion and full-screen cues seem to be effective at improving accuracy, response time, and cognitive load. We also made at-light with good accuracy and response time, implying that, at least in our experiment scenario, people are good at searching for local cues as long as they stand out in some way e.g. animation. In both studies, there were still moving (original target) or full-screen cues (dark, tunnel) that did not perform well, which hints at the complexity of the design space; we cannot blindly trust a single or pair of our design variables. We did not see any cue perform worse than no cue, implying that adding visual cues to teleoperation poses little risk, and can add many benefits.

Our process, choosing design parameters and iteratively exploring implementations through performance and user feedback helped us design our best visual cues (fast target, and framed bounce). This is to say, iterative design processes applied to two parameters yielded useful results. Further, studying multiple parameters at once (our 2x2 design) also revealed interaction effects due to design parameters we did not explicitly study (speed). This leads us to recommend avoiding classic single-variable studies, as much richer results can be achieved in a complicated scenario such as teleoperation.

Our work used a scenario that was close to a complicated, real-world task. While the lights our scenario could even be considered more obvious than other urban search and rescue search targets, the no-cue case proved difficult and the results from the perception literature was still effective. The experiment procedure was also a success as the video monitoring method removed the experimental complications involved with controlling multiple robots, while enabling us to easily swap different scenarios (videos) or trying new cue types. We stress, however, our belief that using real robots to record the video is an important ecological consideration. While our study was still "in-lab," we believe our results further validates the attention literature in "messier" ecological scenarios.

False cues were a core component of our scenario design, mimicking previous work, as well as more realistically portraying how real cues in search and rescue would work. A large risk we anticipated was that false cues would undermine trust in the interfaces, and affect performance. Nowhere, however, did we see the false cue rate impact performance enough to counteract the benefit from cueing. Further, no-cue accuracies were low (66-80%) even in our controlled scenario with spaced-out stimuli. This should embolden robot designers to use modern computer vision algorithms to augment their interfaces even if they are moderately unreliable. This technology can improve robotic teleoperation *right now*.

Table 2. Summary of design parameter effects.

Cue Property	Benefits	Drawbacks
Motion	Higher accuracy	Distracting
Static	-	Poor accuracy, response time, cognitive load
Full-screen	Lower cognitive load	Careful design required for good accuracy and response time
At-light	High accuracy, response time	Increased cognitive load

9. Limitations and Future Work

While we demonstrated that changing our two design parameters could affect operator performance, we did not explore the full continuum of these dimensions. For example, our motion cues were only animated for a short time. Similarly, target was a full-screen cue, but after converging on the light, it was an at-light cue. It may be that different positions on the dimensions of cue-proximity and motion, may have different results and complex interactions. Exploring these and other design cues in the context of urban search and rescue, such as color, animation speed, length, and even non-visual cues such as sound, will help future interfaces for teleoperation.

We introduced miscues into our data, but no statistical results were found. We believe this is due to low numbers (20% of our stimuli were miscues). In response-critical situations where operators must correctly identify regions of interest that need additional resources (e.g. life-saving medical personnel), missing or mistakenly seeing a stimulus may incur great cost. While we focused on performance measures, future research could target these miscues specifically by longer experiments or greater participant numbers that allow greater data to be collected, extending the existing visual attention research on miscues to teleoperation.

Another theme that emerged was that cues that grabbed attention could be perceived as either helpful or distracting, similar to previous work [46]. Our experiment hinted that the positive or negative association may be linked to cognitive load – not task performance – and how a cue could be ignored after it was noticed. This is an important balance to achieve with multiple stimuli on-screen at once (a more realistic scenario than ours), as a cue may be so distracting it distracts from other cued stimuli. Investigating performance with multiple concurrent stimuli and stimuli densities would better illuminate how cues can draw attention away from multiple video feeds.

Finally, this research was in the context of teleoperation, but no robots were actually controlled. Single robot teleoperation remains a challenging open problem that will take much of an operator's cognitive resources, so exploring visual attention while actually controlling a robot is important future work. Moving away from teleoperation, our work may be further generalized by comparing visual search with a still camera and moving target as opposed to our moving camera with still targets.

10. Conclusion

In this paper, we introduced our investigation of the effectiveness of visual attention drawing cues in multi-robot control context. To explore cue proximity and cue motion, we designed and evaluated seven different visual cues through an iterative design process. In our mock-disaster scenario, participants found our search task difficult and our cues useful. Our design parameters had tradeoffs in performance and cognitive load, and our results indicated that full-screen and animated cues can improve accuracy, response time, and cognitive load if they are designed well. Our research provides a baseline for more research to understand cueing visual attention in teleoperation.

11. REFERENCES

1. Richard A. Abrams and Shawn E. Christ. 2003. Motion onset captures attention. *Psychological Science* 14, 5: 427–432. http://doi.org/10.1111/1467-9280.01458

2. Richard A. Abrams and Shawn E. Christ. 2006. Motion onset captures attention : A rejoinder to Franconeri and Simons (2005). *Perception & Psychophysics* 63130, 1: 114–117.

3. Pradeep K. Atrey, M. Anwar Hossain, and Abdulmotaleb El Saddik. 2008. Automatic scheduling of CCTV camera views using a human-centric approach. *IEEE International Conference on Multimedia and Expo*, IEEE, 325–328. http://doi.org/10.1109/ICME.2008.4607437

4. Brian P. Bailey and Shamsi T. Iqbal. 2008. Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Transactions on Computer-Human Interaction* 14, 4: 1–28. http://doi.org/10.1145/1314683.1314689

5. Ali Borji, Ming-Ming Cheng, Huaizu Jiang, and Jia Li. 2015. Salient Object Detection: A Benchmark. *IEEE Transactions on Image Processing* 24, 12: 5706–5722. http://doi.org/10.1109/TIP.2015.2487833

6. B. Bridgeman, D. Hendry, and L. Stark. 1975. Failure to detect displacement of the visual world during saccadic eye movements. *Vision Research* 15, 6: 719–722. http://doi.org/10.1016/0042-6989(75)90290-4

7. Neil D B Bruce, Christopher Catton, and Sasa Janjic. 2016. A Deeper Look at Saliency: Feature Contrast, Semantics, and Beyond. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. http://doi.org/10.1109/CVPR.2016.62

8. Moira Burke, Anthony Hornof, Erik Nilsen, and Nicholas Gorman. 2005. High-Cost Banner Blindness : Ads Increase Perceived Workload , Hinder Visual Search , and Are Forgotten. *ACM Transactions on Computer-Human Interaction* 12, 4: 423–445. http://doi.org/10.1145/1121112.1121116

9. Jessie Y C Chen, Michael J. Barnes, and Michelle Harper-Sciarini. 2011. Supervisory control of multiple robots: Human-performance issues and user-interface design. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews* 41, 4: 435–454. http://doi.org/10.1109/TSMCC.2010.2056682

10. Marvin M. Chun. 2000. Contextual cueing of visual attention. *Trends in Cognitive Sciences* 4, 5: 170–178. http://doi.org/10.1016/S1364-6613(00)01476-5

11. Fiona Donald, Craig Donald, and Andrew Thatcher. 2015. Work exposure and vigilance decrements in closed circuit television surveillance. *Applied ergonomics* 47, January 2016: 220–8. http://doi.org/10.1016/j.apergo.2014.10.001

12. J.L. Drury, J. Scholtz, and H.a. Yanco. 2003. Awareness in human-robot interactions. *IEEE International Conference on Systems, Man and Cybernetics.* 1, October. http://doi.org/10.1109/ICSMC.2003.1243931

13. Steven L Franconeri and Daniel J Simons. 2005. The dynamic events that capture visual attention: A reply to Abrams and Christ (2005). *Perception & psychophysics* 67, 6: 962–6. http://doi.org/Doi 10.3758/Bf03193623

14. Shinji Fukatsu, Yoshifumi Kitamura, Toshihiro Masaki, and Fumio Kishino. 1998. Intuitive control of "bird's eye" overview images for navigation in an enormous virtual environment. *ACM Virtual Reality Software and Technology*, 67–76. http://doi.org/10.1145/293701.293710

15. Dylan F. Glas, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. 2012. Teleoperation of multiple social robots. *IEEE Transactions on Systems, Man, and Cybernetics* 42, 3: 530–544. http://doi.org/10.1109/TSMCA.2011.2164243

16. Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Worload*. 139–183. http://doi.org/10.1016/S0166-4115(08)62386-9

17. Sunao Hashimoto, Akihiko Ishida, Masahiko Inami, and Takeo Igarash. 2011. TouchMe: An Augmented Reality Based Remote Robot Manipulation. *International Conference on Artificial Reality and Telexistence*: 1–6.

18. Eric Horvitz, Andy Jacobs, and David Hovel. 1999. Attention-sensitive alerting. *Uncertainty in artificial intelligence*, 305–313.

19. Eric Horvitz, Carl Kadie, Tim Paek, and David Hovel. 2003. Models of attention in computing and communication. *Communications of the ACM* 46, 3: 52. http://doi.org/10.1145/636772.636798

20. Christina J Howard, Tomasz Troscianko, Iain D Gilchrist, Ardhendu Behera, and David C Hogg. 2009. Searching for threat : factors determining performance during CCTV monitoring. *Security*: 1–7.

21. Denis Kalkofen, Eduardo Veas, Stefanie Zollmann, Markus Steinberger, and Dieter Schmalstieg. 2013. Adaptive ghosted views for Augmented Reality. *IEEE International Symposium on Mixed and Augmented Reality*, October: 1–9. http://doi.org/10.1109/ISMAR.2013.6671758

22. Brenden Keyes, Robert Casey, Holly a Yanco, Bruce a Maxwell, and Yavor Georgiev. 2006. Camera placement and multi-camera fusion for remote robot operation. *Proceedings of the IEEE International Workshop on Safety, Security and Rescue Robotics*: 22–24.

23. Raymond M Klein, W Joseph Macinnes, Raymond M Klein, and W Joseph Macinnes. 1999. Inhibition of Return Is a Foraging Facilitator. 346–352. http://doi.org/10.1111/1467-9280.00166

24. Tomáš Krajník, Vojtěch Vonásek, Daniel Fišer, and Jan Faigl. 2011. AR-Drone as a Platform for Robotic Research and Education. . 172–186. http://doi.org/10.1007/978-3-642-21975-7_16

25. Árni Kristjánsson, Ómar I. Jóhannesson, and Ian M. Thornton. 2014. Common Attentional Constraints in Visual Foraging. *PLoS ONE* 9, 6: e100752. http://doi.org/10.1371/journal.pone.0100752

26. Annica Kristoffersson, Silvia Coradeschi, and Amy Loutfi. 2013. A review of mobile robotic telepresence. *Advances in Human-Computer Interaction* 2013. http://doi.org/10.1155/2013/902316

27. Daniel Labonte, Patrick Boissy, and François Michaud. 2010. Comparative analysis of 3-D robot teleoperation interfaces with novice users. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* 40, 5: 1331–1342. http://doi.org/10.1109/TSMCB.2009.2038357

28. Arien Mack and Irvine Rock. 1999. Inattentional Blindness. *Trends in Cognitive Sciences* 3, 1: 39. http://doi.org/10.1016/S1364-6613(98)01244-3

29. Marcus Mast, Zdeněk Materna, Michal Španěl, et al. 2015. Semi-Autonomous Domestic Service Robots: Evaluation of a User Interface for Remote Manipulation and Navigation With

Focus on Effects of Stereoscopic Display. *International Journal of Social Robotics* 7, 2: 183–202. http://doi.org/10.1007/s12369-014-0266-7

30. D. Scott McCrickard and C.M. Chewar. 2003. Attuning Notification Design to User Goals and Attention Costs. *Communications of the ACM* 46, 3: 67–72.
31. Dan R. Olsen and Stephen Bart Wood. 2004. Fan-out: measuring human control of multiple robots. *Human factors in computing systems*, ACM Press, 231–238. http://doi.org/10.1145/985692.985722
32. Jason Pascoe, Nick Ryan, and David Morse. 2000. Using While Moving: HCI Issues in Fieldwork Environments. *Transactions on Computer-Human Interaction (TOCHI) - Special Issue on Human-Computer Interaction with Mobile Systems* 7, 3: 417–437. http://doi.org/10.1145/355324.355329
33. M. I. Posner, M. J. Nissen, and W. C. Ogden. 1978. Attended and unattended processing modes: The role of set for spatial location. In *Modes of Perceiving and Processing Information*. Psychology Press.
34. Michael I. Posner. 1980. Orienting of attention. *Q.J Exp.Psychol.* 32, 1: 3–25. http://doi.org/10.1080/00335558008248231
35. Z. W. Pylyshyn and R. W. Storm. 1988. Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial vision* 3, 3: 179–197. http://doi.org/10.1163/156856888X00122
36. Sina Radmard, Ajung Moon, and Elizabeth A Croft. 2015. Interface Design and Usability Analysis for a Robotic Telepresence Platform. *Ro-Man 2015*, 6.
37. R. A. Rensink, J. K. O'Regan, and J. J. Clark. 1996. To see or not to see: The need for attention to perceive changes in scenes. *Investigative Ophthalmology and Visual Science* 37, 3: 1–6. http://doi.org/10.1111/j.1467-9280.1997.tb00427.x
38. J. Richer and J.L. Drury. 2006. A video game-based framework for analyzing human-robot interaction: characterizing interface design in real-time interactive multimedia applications. *ACM SIGCHI/SIGART conference on human-robot interaction*: 266–273. Retrieved from http://portal.acm.org/citation.cfm?id=1121287
39. Daniel Saakes, Vipul Choudhary, Daisuke Sakamoto, Masahiko Inami, and Takeo Igarashi. 2013. A teleoperating interface for ground vehicles using autonomous flying cameras. *International Conference on Artificial Reality and Telexistence (ICAT)*, IEEE, 13–19. http://doi.org/10.1109/ICAT.2013.6728900
40. Daisuke Sakamoto, Koichiro Honda, Masahiko Inami, and Takeo Igarashi. 2009. Sketch and run: a stroke-based interface for home robots. *Conference on Human Factors in Computing Systems*: 197–200. http://doi.org/10.1145/1518701.1518733
41. Daniel J. Simons and Christopher F. Chabris. 1999. Gorillas in Our Midst: Sustained Inattentional Blindness for Dynamic Events. *Perception* 28, 9: 1059–1074. http://doi.org/10.1068/p281059
42. Ashish Singh, Stela H. Seo, Yasmeen Hashish, Masayuki Nakane, James E. Young, and Andrea Bunt. 2013. An interface for remote robotic manipulator control that reduces task load and fatigue. *IEEE ROMAN*: 738–743. http://doi.org/10.1109/ROMAN.2013.6628401
43. Matthew J. Stainer, Kenneth C. Scott-Brown, and Benjamin W. Tatler. 2013. Looking for trouble: a description of oculomotor search strategies during live CCTV operation. *Frontiers in human neuroscience* 7, September: 615. http://doi.org/10.3389/fnhum.2013.00615
44. Adrian Stoica, Federico Salvioli, and Caitlin Flowers. 2014. Remote control of quadrotor teams, using hand gestures. *ACM/IEEE international conference on Human-robot interaction*, ACM Press, 296–297. http://doi.org/10.1145/2559636.2559853
45. Peter Tarasewich. 2003. Designing mobile commerce applications. *Communications of the ACM* 46, 12: 57. http://doi.org/10.1145/953460.953489
46. Dan Tasse, Anupriya Ankolekar, and Joshua Hailpern. 2016. Getting Users' Attention in Web Apps in Likable, Minimally Annoying Ways. *Conference on Human Factors in Computing Systems*: 3324–3334. http://doi.org/10.1145/2858036.2858174
47. Wei Chung Teng, Yi Ching Kuo, and Rayi Yanu Tara. 2013. A teleoperation system utilizing saliency-based visual attention. *IEEE International Conference on Systems, Man, and Cybernetics*: 139–144. http://doi.org/10.1109/SMC.2013.31
48. Ian M. Thornton, Heinrich H. Bülthoff, Todd S. Horowitz, Aksel Rynning, and Seong-Whan Lee. 2014. Interactive Multiple Object Tracking (iMOT). *PLoS ONE* 9, 2: e86974. http://doi.org/10.1371/journal.pone.0086974
49. Antonio Torralba, Aude Oliva, Monica S Castelhano, and John M Henderson. 2006. Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological Review* 113, 4: 766–786. http://doi.org/10.1037/0033-295X.113.4.766
50. S. Treue and J. C. Martínez Trujillo. 1999. Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* 399, 6736: 575–579. http://doi.org/10.1038/21176
51. K. VanMarle and B. J. Scholl. 2003. Attentive Tracking of Objects Versus Substances. *Psychological Science* 14, 5: 498–504. http://doi.org/10.1111/1467-9280.03451
52. Eduardo Veas, Erick Mendez, Steven Feiner, et al. 2010. Directing Attention and Influencing Memory with Visual Saliency Modulation. *ACM Conference on Human Factors in Computing Systems*: 1471–1480.
53. Holly A. Yanco and Jill Drury. 2004. Classifying human-robot interaction: An updated taxonomy. *IEEE International Conference on Systems, Man and Cybernetics* 3: 2841–2846. http://doi.org/10.1109/ICSMC.2004.1400763