Reviewer 2:

Summary

The authors set up a simulation system to test the behavior of a UAV operator in a search-and-rescue scenario involving a swarm of UAV's. They conducted an experiment examining the effects of prior knowledge about the missing person, prior knowledge of terrain features, and the distribution of other UAV's in the swarm. EEG and gaze tracking were used to evaluate aspects of the operator's performance.

As a reader, the motivation behind this study was unclear to me, and the results were not clearly presented, so I was not sure what learnings I was meant to take away.

*Create separate section called related work within which have subsections of prior knowledge, swarm perception, cognitive load, and gaze.*

Strengths

+ The Search-and-Rescue problem space is an important area of research with a great potential for real-world impact.

+ The authors seem to have put great effort into setting up a very sophisticated system for running useful experiments, both on the simulation side and on the measurement side.

Weaknesses

- The paper is not motivated well. It is not clear what the experimenters expected to learn from the experiment, and it is not clear why the reported observations are important. For example, what is the significance of the finding that cognitive load and gaze fixation time on an overhead view are affected by terrain knowledge? Is this good or bad?

*Gaze fixation time in the area of interest can be used to provide additional information that a participant may not retain during a briefing [citations supporting additional info should only be provided when needed]. Specifically, gaze fixation can be used to control the onboard camera to provide additional information about the trail.*

*In the conclusion we find that gaze is correlated with prior knowledge and therefore can be used to provide additional trail information in the inset whenever the user is spending time looking at it.*

*We have edited the introduction to add the following information:*

*“… Participant gaze and EEG is recorded during experiments for subsequent analysis. We focus on estimating gaze fixation in an area of interest on the computer display. There can be situations where the operator may have trouble recalling prior information, for instance information regarding location of trails or other geographical peculiarities. In such cases, gaze fixation can be used to judge the operator’s level of prior knowledge and provide them with relevant information. In the conclusion section, we find that gaze fixation on the inset map is correlated with prior knowledge. EEG is recorded to always monitor the cognitive load of the operator during the search mission. A high cognitive load corresponds to the operator struggling to complete a task. If a swarm of autonomous robots are assisting the operator with the mission, they might be programmed to trust the operator less and perform more independent search in cases of high cognitive load.”*

Are the findings suggesting that it is more important to give the teleoperators terrain knowledge ahead of time, or is there may be some suggestion about how to improve graphical interfaces for UAV operators? I did not understand the purpose of this study from the beginning and did not understand the value of the findings at the end.

*Write a paragraph in introduction starting with the* *“The purpose of this study is to quantify the differences in performance, cognition and perception in search and rescue missions as a function of prior knowledge. For field implementation, prior knowledge is then inferred using movement data to then inform a swarm control algorithm where an autonomous swarm searches the terrain in relation to operator prior knowledge.”*

*We have added the following paragraph:*

*“The purpose of this study is to quantify the differences in performance, cognition and perception in search and rescue missions as a function of prior knowledge. For field implementation, prior knowledge is then inferred using movement data to then inform a swarm control algorithm where an autonomous swarm searches the terrain in relation to operator prior knowledge. A reactive autonomous swarm monitoring the operator’s confidence level in real time is central to increasing search performance. The confidence level is measured in the form of prior knowledge and will be helpful in determining the degree of autonomy of the swarm.”*

- Novelty of this work is questionable. The authors cite studies such as [12, 13] which have already demonstrated the importance of prior knowledge of the person and terrain, so what does the current study contribute beyond this? Maybe in some way this is meant as a replication study, but this is not stated in the paper. Similarly with the UAV positioning, the authors cite [16], so what does this paper contribute beyond that?

*Create and contrast with related work before writing what is done in this study.*

- Soundness of the findings are questionable due to what appear to be oversimplifications.

-- User modeling

The modeling of the movement of the user appears not to be sophisticated. To my understanding there is already substantial work modeling how people move when lost according to terrain, availability of water, and so on. This model seems to only use random placement.

*Say it is a simple model used to highlight the effects of prior knowledge and not the effect of lost person behavior (cite the random walker).*

*“…1, 2, and 3 m/s respectively. The lost person behavior was made simple on purpose because the purpose of this study is to highlight the effects of prior knowledge. The position of the …”*

Also, the walking speeds of 1,2,3 m/s seem unusual for a person lost in the wilderness - 3 m/s is a running speed. It is also unclear why 3 m/s applied over 36 hours would only result in a radius of 1.08km. If these numbers are grounded in any research, the authors should cite it. As presented, it seems like a simplified game, and it is unclear to what extent these results will apply to real situations.

*Clarify the calculations of random walk.*

*We have added text to reflect the following:*

*“… may be found based on their walking speed. The radii of the POAs were set as the distance travelled by a random walker walking at speeds 1, 2 and 3 m/s in 36 hours. The average distance was calculated as (Papoulis, 1991) which translated to 0.36 km, 0.72 km, and 1.08 km from the PLS. The position of the teleoperated …”*

Papoulis, A. "Probability, random variables, stochastic processes (MGH, 1991)." (1991).

-- Swarm distribution

The swarm distribution part of the experiment was very confusing to me. As I understood the paper, the swarm was either evenly distributed or clustered at a random point unrelated to the missing person's location. This seems meaningless to me if it is not based on actual swarm search dynamics. The finding that this influenced mission performance seems like it is probably spurious.

*A swarm searching for a missing person maybe doing as a cluster or as a independent search. In either of these cases, the swarm will not have any additional knowledge beyond the operator. In this context, we wanted to understand how different search strategies do --- searching together vs independently --- affect operator behavior, which in turn could be used to regulate such behavior.*

*Create a 2d heatmap of the operator as a function of swarm distribution: Every trajectory when the swarm was clustered. Another heatmap for when the swarm was uniform.*

*Subtract start position, take angle b/w start location and end location. Theta is angle b/w x axis and the vector that joins the start and the end position*

*Create another plot: 2D distance to the center of the swarm as a function of time. One of clustered and another for uniform.*

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |

-- Terrain knowledge

The movement logic used for the missing person is not adequately described, so I was unable to understand why terrain knowledge would be relevant. Certainly, the challenge of mentally visualizing the locations of trails not shown on the display would be difficult for operators, but it was unclear to me whether knowledge of trail locations would help the search effort in this scenario.

*Add more details on how we modeled the missing person behavior.*

*Details were added to clarify the missing person behavior.*

*“… red smoke that dissipated with distance. The distance travelled by the missing person from point last seen is approximated as the distance travelled by a random walker in 36 hours at 2 m/s along a trail. The missing person was placed randomly within a radius of 100 m from the intersection of a trail and a point between the center and outer POA. PLS locations were chosen to be on trails such that the trails extend well beyond the largest POA radius (corresponding to 3 m/s random walk) in at least two directions.”*

Reviewer 4:

Comments to the author

======================

The reviewed paper aims to evaluate how knowledge of an incident’s context affects a drone operator’s effectiveness in a missing person search as well as their cognitive load and behavior. A controlled experiment is run where the subject incrementally has access to prior knowledge of the missing person, terrain, and the distribution of a swarm of drones. During each trial, data is recorded on the user’s operating behavior including time to find the person and turn rate. There are also detailed physiological measures that aim to measure the cognitive workload and perceptive dwell time.

The overall methodology and presentation of results are sound and it’s clear that substantial effort was put into this work. I thought the approach towards collecting cognitive workload was thorough in its use of EEG collection and survey responses. There is a good level of detail presenting differences in the set of results across experimental conditions. The results could have been improved with more presentation of the actual trajectories flown and if any specific behaviors are elucidated.

*Add all 2D top view trajectories in supporting information. Add 2D view in main manuscript as well.*

*We have added a top view plot of the trajectories for all participants, with the POAs marked for each condition.*

A graph with green circles

Description automatically generated

The biggest weaknesses that this paper presents is the divergence in applicability towards real world problems as presented in the authors fundamental research questions. The first research question asks whether more knowledge of context leads to better search performance. Any practitioner of search and rescue can give a resounding “YES!” to this question, which is why this information is always contained in a standard briefing, as would be done in this type of search mission. While large scale disaster response incidents may lead to the operator having more uncertainty about the terrain, they would still leverage geographical knowledge to the best of their ability.

*Elaborate in the introduction, the amount of prior knowledge and cite instances where it may be misinterpreted. Slow moving vs fast moving, injury may happen as they were missing and so a dynamic inference of prior knowledge maybe useful in controlling an autonomous swarm.*

The authors also improperly calculate their predicted missing person radii given the reported 36 hours of disappearance to a maximum distance of 1.08km.

*Clarify the calculations of random walk.*

The second research question aims to consider operator responses to other drones operating in the area, which may have some merit. However, there is no information provided to the operator as to how the other drones are coordinating their search operations. Any incident of this nature would have all aircraft operators being coordinated through a central Air Operations supervisor to define airspace delineations and responsibilities. Without any knowledge of the other drones’ operations these assets seem to act as distractors, and I don’t see how operator performance is relevant.

*Add focus on non-verbal cues. In LSTM, switch on the swarm control closed loop after X amounts of seconds based in trajectory analysis.*

Finally, the third research question considers how prior knowledge affects cognitive factors. The approach towards answering this question is thorough although other relevant works that address similar questions are not sufficiently referenced (E.g. Britton et al. “Effects of prior knowledge on use of cognitive capacity in three complex cognitive tasks,” 1982, Innes et al., “The effects of increased visual information on cognitive workload in a helicopter simulation,” 2021.)

*Cite these papers in the introduction and contextualize with respect to this study.*

While measuring cognitive load is a substantial area of ongoing research, unfortunately the context around which they aim to evaluate it (availability of prior knowledge) does not present any fundamental contributions.

*Cite literature that cognitive load is affected by performance, show correlation b/w performance and cognitive load, calc correlation b/w cognitive load and speed and turn rate.*

*Remove cognitive load from stats table.*

*The swarm leave the operator if the 3s cognitive load is above a certain threshold.*

Reviewer 5

Summary: This paper investigates how prior knowledge in a missing person’s speed, prior terrain knowledge, and distribution of UAVs in a search area influences operators’ performance, workload, and actions in a search and rescue teleoperation task. The authors conclude the following:

- prior knowledge of a missing person’s walking speed improves performance and reduces turn rate

- prior terrain knowledge increases cognitive load and increases gaze fixations on an inset overhead view

- the presence of clustered UAV swarms improves performance and increases cognitive load across the task, with reduced cognitive load for the first segment of the task

Comments: This work was well written and easy to read. The work contributes to teleoperation in search and rescue situations by investigating the effects of prior knowledge of the terrain, missing person’s speed, and the presence of UAV swarms on important metrics for teleoperation including performance, workload, cognitive load, gaze fixation, and users’ actions. I recommend the authors clarify the following:

How did the authors determine that they needed 20 participants? If they conducted a power analysis that should be reported in the paper.

*We will alter the p-value based on post hoc power analysis.*

*“An a priori power analysis was conducted using G\*Power for sample size estimation for repeated measures within factors experiment setup. The effect size in was assumed to be 0.5, with a significance criterion of α = .01 and power = .95, the minimum sample size needed with this effect size is 9. Thus, our sample size of 20 participants (16 males and 4 females, aged 24 ± 4 years) is adequate for our study. The participants were recruited through flyers and email announcements. Candidates were excluded …"*

There are potentially confounds in the performance analyses as the upper bounds of the performance metric seems to have been included in the analysis. The 600 second upper bounds is assigned to all participants who did not find the missing person within the allotted time. If those participants were to have unbounded time, their performance could be anywhere from 601+ seconds, however, they are all treated the same. Unfinished runs should be treated as missing data as we do not have an accurate measurement of the participants’ performance.

*Do both. For GLMM kill the trials but report with 600s in supplementary.*

There are potentially confounds in the UAV distribution analyses. It is mentioned in the introduction that the intent on using the drone cluster was “to indicate the possibility of useful information”, however, the clusters were randomly placed and “had no relation to the actual location of the missing person”. The random placement seems to go against the intended benefits, potentially misdirecting participants to focus on random regions. Since the random clusters could only appear in the outer two POAs, I wonder if we would see different results if the uniformly distributed UAVs were also only in the outer two POAs.

*Discuss possible limitations.*

I also wonder if in conditions where prior knowledge was present, drone clusters or distributions in the corresponding regions would help improve performance over the random placements used in this study.

*Discuss this as well.*

*We have added text in the discussion section as follows:*

*“Placement of drone clusters were purposefully made to be randomly placed between the middle and outer POAs because in this study our principal goal is to observe the effect of prior knowledge on search performance with a secondary goal of influencing the operator behavior using drone clusters. In the briefing for cases where the prior knowledge about a human is present, we emphasize that the missing person may be somewhere between the middle and outer POA. Mission performance was strongly influenced…”*

*“If the swarm cluster was placed in the general vicinity of the missing person, the results probably would have diluted the effects of our primary goal. Terrain knowledge did not lower the …”*

Participants reported that they did not find the placement of the UAVs beneficial (Table 1), this should be further discussed along with suggestions to improve this visualization.

*Discuss this.*

*We have added another paragraph in the discussion section as follows:*

*Participants knew that the swarm was searching for the missing person. It is possible that they may have responded differently if instead the swarm was shown to have found the participant and the job of the participant was only to verify. This setup, would have also raised the trust of the participant. However this was not the case, and indeed the participants reported that once they realized that once the swarm is not hovering over the missing person, they stopped going after it.*

*This supports an HSI strategy where the participant independently searches for the missing person in presence of the swarm.*

*“The presence of other UAVs did not seem to affect the overall search behavior of the participants as reported in table 1. This outcome is as expected, because the UAV swarm serves no real purpose other than to distract the participants. Four of the eight conditions have uniformly distributed UAV swarm, and this is ignored by them. In the other four conditions where swarms are clustered at random locations potentially could have been distracting. In a separate questionnaire, a few participants mentioned that they initially searched near the swarm cluster but abandoned it for subsequent conditions.”*

It is unclear why the observation window of 1-25 seconds was selected for further analysis of cognitive load, when Section III D mentions that it levels off after five seconds. This subset of analyses is also missing the effects of swarm distribution.

*Cognitive load is used for closing the loop*

Is dwell time on the inset necessarily a positive effect? A further discussion of the negative effects of participants shifting their focus away from identifying the missing person to focus on the overhead map should be included in Section III G. I recommend the authors also offer suggestions on how to improve the user interface to keep participants focused on the area where they can see missing people.

*Add discussion*

Statistical analyses of the NASA TLX in Section III E are not reported. I don’t see how the claim that “participants felt that they had to work somewhat hard to accomplish the task for conditions where they did not have missing person prior knowledge” can be made, especially since all subscales, except for performance, in the table report means below half, indicating low workload.

*Perform and put stats.*

Typically, an overall score is calculated for the TLX, the authors should report the scores and analyses of these scores even if they are not significant. Statistical analyses should back up all of the authors' claims.

*Find out how to calculate overall score and report it.*

*NASA TLX is actually a 2 step process (https://ntrs.nasa.gov/api/citations/20000021488/downloads/20000021488.pdf). In the first step, we figure out the weighting of each subscale (mental, physical, …), where the participant themselves put the weights and in the second step we ask the participant to fill out the actual ratings form. We haven’t done the first part. However, I have found a paper which mentions that in many cases just reporting the individual subscales are acceptable.*

*The other common variation is to analyze subscale ratings instead of generating a single overall workload score. This was done in at least 40 of the studies I reviewed. In addition, individual subscale analyses were performed in addition to assessing overall workload in nearly 20% of the studies. Both of these approaches demonstrate one of the continuing strengths of the scale; the diagnostic value of the component subscales. The component ratings can help designers pinpoint the source of a workload or performance problem.*

*Hart, Sandra G. NASA-TASK LOAD INDEX (NASA-TLX); 20 YEARS LATER*

If the authors are using the standard NASA TLX scale, the claim “participants generally rated their performance as good” seems incorrect as all performance responses were past the half mark, trending towards the “Failure” (20) end of responses. If non-standard scales are used, they should be reported in the paper.

*Say that we are using a non-standard scale for TLX.*

In Section IV, results might be more clearly discussed if they were broken down in regards to the original three research questions. Further discussions on how the results related to findings or contribute to related literature would also be beneficial.

*Yes, divide it into sections.*

Typos and small suggestions:

Table 1 header “Questions about Virtual Reality” would more accurately be “Questions about Virtual Environment”. There is only one VR question asked, and based on the experimental setup, this study was not conducted in VR.

*OK*

Table 2 header “Fraction of Dwell Time” would be more clear if renamed to “Fraction of Dwell Time on Inset”

*Ok*

In Figure 5. Graphs A, C, and D it is not easy to follow the connected lines to see values for each participant. I would recommend removing them to make the graphs clearer.

*Ok*

Participant gaze is discussed in Section III B, when it has its own subsection: Section III G.

Ok

Section III C. should reference Figure 5a, not 4a.

Ok

Section IV, the authors list “mental workload”, based on their results I believe they mean “cognitive workload” since statistical analyses of the mental workload subscale of the NASA TLX were not reported.

Ok

Suggested related work:

Ewers, J. H., Anderson, D., & Thomson, D. (2023). Optimal path planning using psychological profiling in drone;assisted missing person search. Advanced Control for Applications: Engineering and Industrial Systems, 5(4), e167.

Kashino, Z. (2020). An Adaptive Approach to Optimal Sparse Mobile-Target Search Planning Using Heterogeneous Agents. University of Toronto (Canada). (Doctoral dissertation).

*Cite them*