Assessing the Impact of Autonomy and Overconfidence in UAV First-Person View Training

Songpo Li*, Mary L. Cummings, Benjamin Welton

Department of Electrical and Computer Engineering at Duke University, Durham, NC 27708 USA

Abstract

With the rapid rise in unmanned aerial vehicles (UAVs) for military and civil first-person applications like infrastructure inspection, there is an increased need for skilled UAV operators. However, research on effective training of UAV pilots has not kept pace with the demand. How much autonomy should be onboard, how much training, and how much control humans should have are still points of debate. To help fill this gap, this paper examines how different training programs and levels of control autonomy affect training outcomes for people operating a UAV in inspection tasks with high onboard autonomy. Results revealed a cost-benefit trade space in that those top performers with both lower-level teleoperation and higher-level supervisory control training could achieve the best performance, but with higher variability, as compare to those who received just supervisory control training. Another important finding was that those trainees who were overconfident were more likely to spend too much time micro-controlling the UAV, and also 15 times more likely to crash. Given that commercial UAV licensing is expected to significantly increase in the next few years, these results suggest more work is needed to determine how to mitigate overconfidence bias both through training and design.

Keywords: supervisory control, drone, UAV, pilot training

1. Introduction

The adoption of unmanned aerial vehicles (UAVs) is growing rapidly across diverse military and civil applications due to their mobility, standoff capability, and ease of deployment [1, 2]. Typical UAV uses include many military applications [3], as well as civil applications, including search and

*Corresponding author
Email address: songpo.li@duke.edu (Songpo Li)

rescue [4, 5, 6, 7], remote sensing [8, 9, 10], construction and infrastructure inspection [11, 12], and precision agriculture [13, 14]. There has been considerable recent investment in and growth of UAVs. It is expected that UAVs, aka UAVs, will grow to more than $82.1 billion in annual revenue by 2025 [15]. The FAA has estimated that UAVs will number more than $2.4 million in 2022 [16].

Current UAV operations require intense human supervision [17, 18], and operator error has been one of the major causes of UAV incidents [19, 20, 21]. Previous research has studied how automation levels affect an operator's performance and how autonomy can be dynamically adapted based on mission and operator demands [22, 23]. This research shows that the insertion of higher levels of automation, meant to reduce workload, may not always lead to improved operator performance. The FAA predicts the number of needed commercial UAV pilots will be 350,000 by 2023 [17]. Given this need as well as the excessive amount of operator errors, a fundamental question is how to effectively train UAV pilots, as well as how technology could be improved to assist pilots in training and operation.

Previous research on UAV pilot training has primarily focused on training augmentation with improved training devices and evaluation, as well as pre-screening [24, 25, 26]. UAV simulators have demonstrated some effectiveness in the initial training of pilots [27]. UAV motion cues have been used to increase a simulator's realism and training outcomes [28]. Another study introduced a chase viewpoint from the rear of the UAV to augment an operator's situation awareness in cluttered environments [29]. Functional Near-Infrared Spectroscopy (fNIRS), which measures brain blood oxygenation, was used to estimate UAV pilots' cognitive workload and situational awareness during various flight tasks [30], with mixed results. Researchers in [17] developed a methodology based on fuzzy logic to fuse different metrics, including performance score and cooperation, to assess overall skill achievement. Addition of autonomy assistance can relieve an operator's workload and improve performance. Previous research has studied how automation levels affect an operator‘s performance and how autonomy can be dynamically adapted based on mission and operator demands [22, 23]. This research shows that the insertion of higher levels of automation may not always lead to expected outcomes.

While this previous research focused on aiding or measuring training, very little work has focused on the type and duration of UAV training needed to achieve minimum proficiency, which is of significant interest given the growing commercial industry. The US Air Force has the most experience training UAV pilots, and it prefers to have its cadre of UAV pilots first learn to fly conventional

aircraft before flying UAVs. However, recently due to a pilot shortage, the Air Force allowed some UAV operators to skip learning to fly a plane before learning to remotely operate UAVs. Those operators who learned to command UAVs with no piloting experience often performed similarly as those pilots who transitioned from flying actual aircraft to flying UAVs [31]. So, it is unclear how much time spent learning to manually fly as opposed to the more click-and-point method of flying UAVs (called supervisory control) affects overall proficiency.

To address the limited research in this area, experiments were conducted in an indoor motion-capture facility to evaluate the hypotheses that the addition of lower-level manual control training would enhance overall supervisory control performance. This experiment, focused on training and described in the next section, is an extension of our previous research examining how control interfaces with different levels of autonomy affect performance in UAV operations [32].

In addition to exploring whether training with both lower- and higher-levels of autonomy could produce superior performance as compared to just higher-level autonomy training, we also wanted to explore the hypothesis that mixed training could engender too much trust in the onboard autonomy. In our previous study [32], some operators with both forms of training inappropriately used waypoint control for navigating tight space, when they should have been using nudge control. Their abuse of the onboard autonomy could potentially led to more crashes. At that time, it was not clear whether this was a sign of overtrust of the onboard flight control system. In this experiment, one unexpected set of observations was a disconnect between participants' ratings of their own performances and the actual physical outcomes of each experiment. We were surprised that many people reported good performance despite knowing that they crashed. This further led us to investigate, post hoc, whether operators' overconfidence in their own abilities could influence the outcomes.

Previous research has shown that the degree of an operator's trust in autonomy is a critical element of the joint human-autonomy system performance, as is their self-confidence [33]. Human operators that are overconfident in their own abilities as well as that of the onboard autonomy are likely to misuse (over-rely) the assistance provided by the system [34, 35]. Alternatively, operators who do not trust the autonomy or that lack self-confidence may under-rely on the system and fall short in achieving their objectives [36, 37]. Thus, in addition to determining any performance differences, we added exploring if and how trust and self-confidence would affect outcomes. Over-confidence in one's piloting ability has long been an issue in conventional aviation accidents, and UAV operations can also suffer from such issues [38].

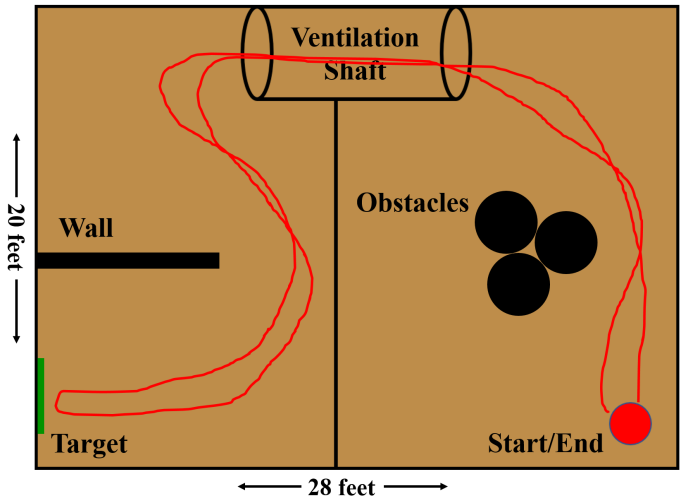


Figure 1: Map of the testing environment

2. Method and Experiment

2.1. Testing Environment

Participants were told that their mission was to fly a UAV through a partially-collapsed enclosed indoor environment due to an earthquake, because it was unsafe for humans to enter, much like the Fukushima scenario. Their primary goal was to navigate the UAV to a control panel and read information from it, depicted as the target in Fig. 1. Each UAV took off from the start point and could only reach the control panel in the left room through a ventilation shaft. After reading the information from the panel, the UAV had to return to the takeoff position through the shaft. The red curved line in Fig. 1 demonstrates a typical UAV flight trajectory. Participants were given a map of the environment, which represented a known map of the building generated before the earthquake. Obstacles were placed in the environment to represent uncertainty in the environment, and the locations of the obstacles were not known to the operator. The operator was notified before the flight that there could be unknown obstacles in the environment and was encouraged to judiciously use the onboard camera to explore the environment to avoid them.

2.2. Application Interface

The UAV could be flown in either through teleoperation, called manual control (MC) or they used a supervisory control (SC) mode. Each mode had a similar, but different control interface.

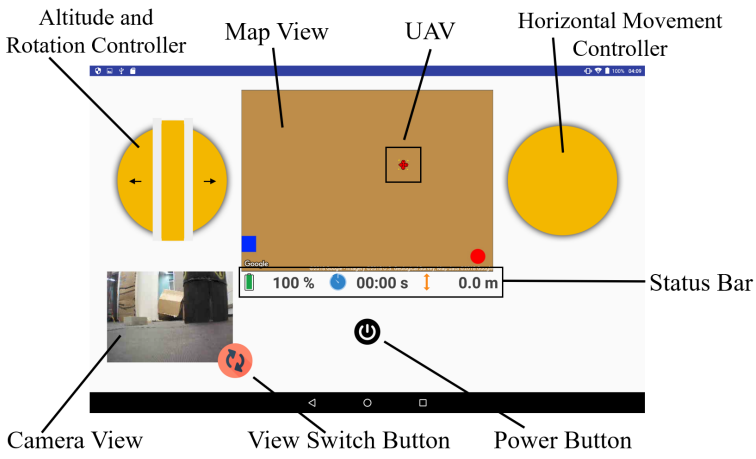


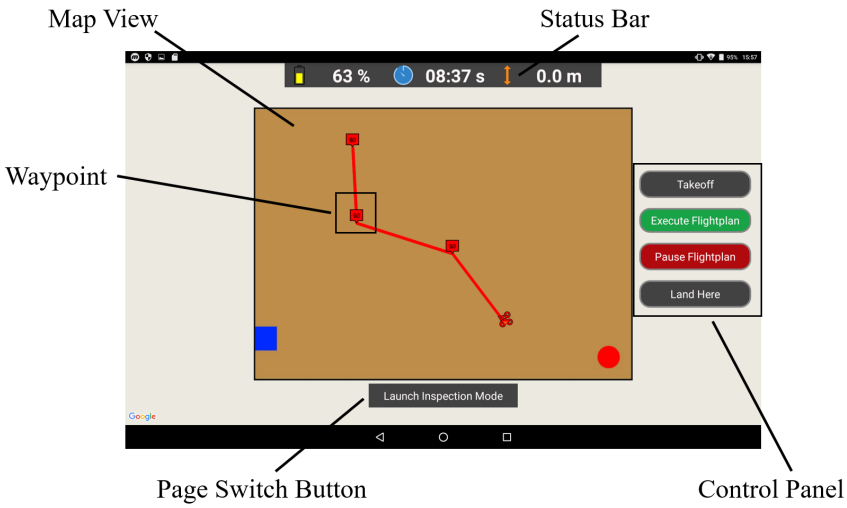
Figure 2: The interface for the manual control mode

The two control modes were implemented on a tablet (Lenovo Tab 2 A10-70) using an open-source software platform called Paparazzi on Parrot AR 2.0 UAVs. A motion capture tracking system consisted of 24 Vicon Vero cameras and 2 Vicon Vantage cameras, which tracked the UAV inside the experiment environment.

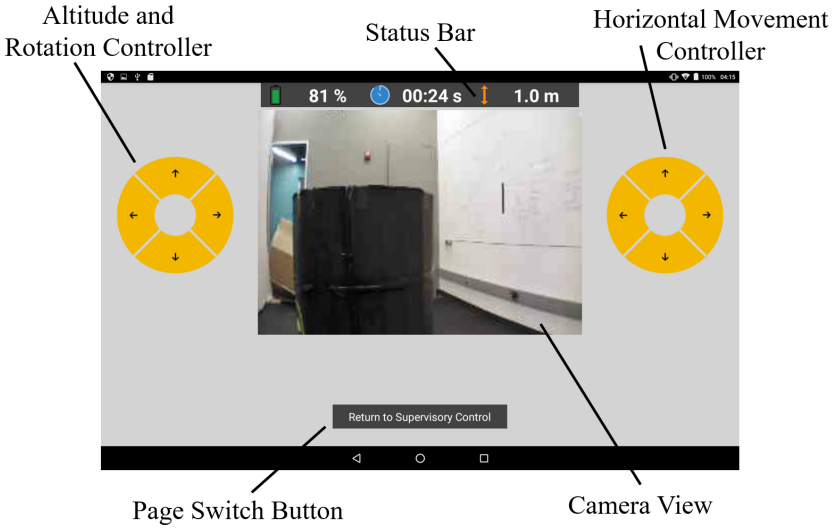
In the MC mode, an operator manually controls the movement of rotation, altitude changes, and horizontal motion in any direction (Fig. 2). There are five major components on the interface: a map view, a camera view, a status bar, a power button, and two joysticks. The location of the UAV is shown on the map in real-time with a red UAV icon, and an arrow on the icon indicates the UAV's heading. The operator navigates the UAV using the two joysticks. With the left joystick, the middle bar adjusts the UAV's altitude by sliding up or down, and the two outer buttons rotate the UAV to change its heading. The middle bar also enables take-off or landing by sliding and then holding it. The right joystick enables directional translation motion (i.e., left or right) at the same altitude.

The status bar summarizes the UAV's battery status, flight time since taking off, and the UAV's altitude. The map view gives the operator global situation awareness, while the operator obtains the local environment information surrounding the UAV through a live camera stream shown in the camera view. The operator can toggle between the two views using the switch view button.

In the SC mode, the UAV has a higher level of automation, where operators provide a goal in



(a)



(b)

Figure 3: The interfaces for the supervisory control mode. (a) The waypoint control, (b) The nudge control.

terms of a waypoint and the onboard autonomy computes and flies the correct trajectory to this point. The operator specifies the UAV's altitude and sets waypoints to build an overall flight plan by assigning ordered waypoints on the map and then commanding the UAV through the execution button (Fig. 3a). The interface also allows for automatic takeoff and landing, and the operator can suspend and revise the flight plan during execution.

The operator can check the surroundings of the UAV by switching to the nudge control, which provides the same camera view as in the MC mode. Switching to the nudge control(Fig. 3b) automatically suspends the execution of the flight plan and causes the vehicle to hover in place. Two joysticks are present on the nudge control page for performing fine positional adjustment, termed nudge control [39]. For example, clicking the left-rotation button will rotate the UAV in small degrees to the left. The two joysticks on the SC nudge control interface are not designed for continuously flying the UAV but only for micro-adjustments.

The main difference between nudge control in the SC interface and MC control is that MC control gives operators rate control, typical of teleoperated systems, while nudge control only makes small position changes. This allows people with manual control to move with more agility as well as faster, while nudge control takes longer to do the same maneuver, but leads to fewer crashes [39].

2.3. Subject Training

Two training programs, manual control plus supervisory control and just supervisory control, were investigated to examine the role of advanced autonomy and required skill sets. Each program consisted of six training modules and a full practice module. The six basic modules were 1) UAV basics, 2) App interface training, 3) Takeoff & landing, 4) General navigation, 5) Camera operation, and 6) Emergency handling, summarized in Table 1[1]. Each module included tutorial slides, followed by quizzes for assessing the learning efficacy before moving to the next module. In addition, modules 3-5 provided hands-on practice, which allowed trainees to practice the particular skill introduced in the tutorial. All training modules were designed to be self-paced, where trainees decided when to move to the new module or operational practice.

[1]The six training modules can be found at https://hal.pratt.duke.edu/UAV-piloting-research

Table 1: Summary of the basic six training modules

Module		Description	Content	
			Slide & Quiz	Hands-on
Training Modules	1	Briefly explain basic aerodynamics of UAVs	Yes	None
	2	Introduce the corresponding control mode and interface	Yes	None
	3	Explain how to take off and land	Yes	Yes
	4	Explain how to navigate a UAV	Yes	Yes
	5	Explain how to adjust the camera view by controlling the UAV	Yes	Yes
	6	Provide general advice and steps for possible emergency handling	Yes	None
Full practice		A full flight practice to ensure readiness for mission	None	Yes

Table 2: Experimental treatments

	Group 1	Group 2	Group 3
Training	MC-SC/SC-MC	SC	SC+
Training time allotment (mins)	120	82	120
Testing	SC	SC	SC
Testing Time (mins)	15 + 15 (optional) + 15 (optional)		

2.4. Experiment Procedure

A demographic survey was filled out by each participant after signing the IRB-approved informed consent form. Participants were paid $25/hour with Amazon gift cards. The best performer in each group who completed the initial testing mission with the least time won an additional $100 gift card, which provided an incentive for better performance.

Participants were randomly assigned to three different training groups (Table 2). Group 1 had manual control (MC) training and supervisory control (SC) training. Group 1 was split into 1a and 1b where the order of training was reversed (MC+SC vs. SC+MC), as we wanted to check for any order effect in the presentation of training material. People in Group 1 completed two training syllabi as outlined in Table 1, except they only experienced module 1 once. Thus, people in Group 1 practiced MC and SC in all hands-on tasks as outlined in Table 1, but only completed the tests with SC.

Group 2 was only trained to use the SC mode, which resulted in reduced training time. Group 3 also had just SC training, but in order to determine if time in training was a significant factor, Group 3 had an additional hands-on practice that allowed them to match the overall time allotment of Group 1. This group is referred to as SC+. Even though the three groups were trained differently, all three were tested on the SC mode only in the post-training missions. These three groups allowed us to determine whether the addition of manual control training adds any additional value as compared to just supervisory control training.

As indicated in Table 1, all participants completed a full mission practice using their assigned control mode or modes at the end of training, just prior to the final mission tests. There was a 20-minute checkride at the end of training to ensure participants mastered the requisite skills of drone navigation and camera operations. Furthermore, all participants practiced the hands-on training and took the checkride in environments different from the final mission test environment.

The final mission testing session consisted of three independent testing trials and it took up to three hours to complete a single test session. The first one was mandatory, and trials 2 and 3 were voluntary based on subjects' willingness and time. On average, for those participants who elected to finish multiple tests, they took an approximate five-minute break between test sessions. Participants were told that their objective was to navigate the UAV to a control panel, read out loud information from it, and then return to the starting location. After the first test, participants filled out a post-experiment survey to capture their subjective opinions on the training, system, and their skills. Then, based on the time and participants' willingness, they could elect to participate in up to two additional rounds of testing (so there were three possible trials). Changes were made to the testing environment between each trial so that the environment appeared different.

3. Results

Forty-six participants participated in the experiments, and data from four participants was excluded from analysis due to system failures or participants' inability to pass the final checkride. Of the 42 that completed the first trial, 39 completed two sessions, and 33 participants completed all three test trials. Seven people from Group 1 dropped by Trial 3, and one each from Groups 2 and 3. Because training for Groups 1 took longer than other groups, many participants in Group 1 ran out of time to attempt additional trials. Mission failure occurred when the participant crashed the UAV or it ran out of battery power before accomplishing the mission. For all statistical

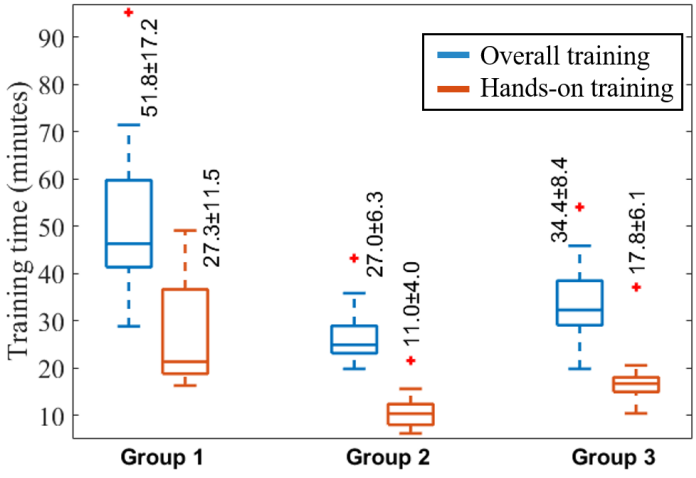


Figure 4: Boxplots of participants' overall training times (blue) and hands-on practice times (orange), with means and standard deviations.

analyses, a significance level of $\alpha = 0.05$ was used. All ANOVA models met required normality and homogeneity of variance assumptions. There were no statistical differences detected between the participants in Groups 1a and 1b for any metric described, so the results are presented simply as Group 1.

3.1. Participants

The 42 participants' ages ranged from 19-64 yrs, with a mean of 25.8 yrs and a standard deviation of 9.7 yrs. All three groups had 14 participants. This number was determined based on a previous similar experiment [32], as well as practical limitations like motion capture room scheduling constraints. Out of the 42 valid participants, 14 were female, and 28 were male. Thirteen participants reported experience in UAV operations with a joystick controller or using a touchscreen device (e.g., phone, tablet). One subject from Group 2 had UAV operation experience above 50 hrs based on self-estimation. The other 12 subjects had an average of 5.38 hrs of UAV experience (std dev = 5.63 hrs).

Table 3: Details on the success rates for the three groups in the first trial.

	Group 1	Group 2	Group 3
Success count	4	5	7
Failure count	10	9	7
Total	14	14	14
Success rate	0.29	0.36	0.5
Success completion times (mean ± st. dev.)	6.22±1.0	6.21±0.93	6.74±1.21

3.2. Training Times

The overall training time participants spent on the self-paced training modules and flight practice is shown in Fig. 4, as well as the time spent on hands-on flight practice. The difference is the time spent on the module training. Given that Groups 1 and 3 could spend up to 120 minutes in total training, and Group 2 up to 82 minutes, it is clear from Fig. 4 that people quickly went through both the self-paced tutorials as well as the flying elements. The groups could each spend up to 73, 52, and 78 minutes respectively in the fight portion, but they were allowed to stop once they felt their performance was satisfactory.

While training times were designed to be the same for Groups 1 and 3, a one-way ANOVA revealed a significant difference across the three groups ($F(2,39) = 16.8$, $p < .001$). In most cases, people elected to use half (or less) of the training time allotted to them, so these training times reflect individual perceptions of how much training participants thought they needed. Of particular note, people in Group 3 who had the opportunity to fly the UAV for up to 78 minutes, on average, only flew for 18 minutes (23% of allotted flight time).

3.3. Results from the First Trial

Table 3 summarizes the success rates of the groups in the first trial. Group 3 performed the best with 50% of participants accomplishing the mission, and Group 2 had the second-highest success rate (36%). The success rates across the groups were not statistically different. There were no significant associations between the training times and whether people crashed. Among the 16 successful trials, five were from participants with drone flying experience, and the other 11 were from those without previous drone experience. The five participants with drone experience used

significantly less time (5.54±0.86 mins) to complete the flight task than the other 11 participants (6.85±0.95 mins). Using this experience as a blocking factor in a two-way ANOVA with group as the second factor, there was no statistical difference between groups for completion time, but there was a significant effect for experience ($F(1,12) = 5.04$, $p = .044$).

Among the 26 failures, five were time-out failures, i.e., people crashed because the battery was depleted, but this group was too small for further analysis. There were 18 participants who played games as a hobby, and there was no statistical association between gaming experience and whether people crashed (Cramer's $V = .014$, $p = .927$).

To study the pilots' confidence, participants were asked to rate on a Likert scale (1 = strongly disagree and 5 = strongly agree) how confident they felt about using waypoint and nudge control, as well as how much they trusted the onboard autonomy for waypoint navigation and nudge control. They were also asked which of these modes they preferred to use while navigating in general and negotiating tight spaces.

Using a one-way ANOVA, there were no statistical differences across the groups in terms of operator self-reported confidence in using either waypoint ($M = 3.9 \pm 1.1$) or nudge control (3.4 ± 1.3). Participants were slightly more confident in using waypoint control than they were in using nudge control. To determine if this may have been a factor in crashing, people's waypoint confidence scores were compared between those who crashed and those who succeeded, but there was no statistical difference.

While there were no differences in confidence between the two modes, this was not the case for trust. There was no statistical difference between groups for trust in nudge control ($M = 4.0 \pm 1.1$), but there was a difference for trust in waypoint control ($F(2,39)=3.28$, $p = .048$). People in Group 1 who experienced both types of training reported a mean of 2.9 ± 1.1, while the other two groups with just supervisory control training reported an average trust rating of waypoint control of 3.8 ± 1.1). When the trust in waypoint ratings were compared between those who succeeded and failed, there was also a significant result ($F(1,40) = 5.17$, $p = .028$). People who crashed had lower trust ratings of waypoint control (M 3.2 ± 1.1) as compared to people who succeeded ($M = 4.0 \pm 1.0$).

One of the significant benefits of waypoint control, a higher level of autonomy, is workload reduction. If workload is too high, it can degrade human performance [40]. To examine workload, which is a subjective metric in that each individual perceives the amount he or she is working differently [40], a one-way ANOVA was run with successful people across the three training groups

with workload reported on a 5-pt. Likert scale (1 = extremely low, 5 = extremely high). There was a statistical difference between the groups (F(2,39) = 3.5, p = .040). People who experienced manual control training in addition to supervisory control training, not surprisingly, reported the lowest workload (M = 2.0±.7). People who trained with the short supervisory control program (Group 2) reported the highest workload (M = 2.7±.5), while people with just SC who were given more time reported slightly less workload (M = 2.6±1.0). In addition, there was a significant Spearman rho correlation between overall training time and workload (coefficient = .-39, p = .016). These results mean that, generally, those people who trained more experienced less workload. However, there was no direct association between workload and whether people crashed.

3.4. Results from Multiple Trials

Because people could elect to fly up to two more times in slightly modified environments, we could examine performance over time. There were a total of 114 completed sessions with 62 successful attempts (53%). Such high incomplete rates are typical for such complex indoor trials [32]. Because training for Group 1 was longer (Fig. 4), these people were less likely to complete all three trials. There was a statistically significant Cramer's V association between the different trials and the number of successes ($p = 0.042$), meaning that as trials progressed, there were fewer crashes.

Fig. 5 shows the box plots of the completion times for the three trials. Across the participants, Group 2 had the lowest average completion time in Trial 1 (Table 4). In Trial 2, Group 1 had the lowest average completion time. In Trial 3, Group 1 had the shortest average completion time of 3.64 minutes, with a 41% reduction from the Trial 1, which was the biggest time reduction for any group. For completion times, using a two-way ANOVA for groups and trials, there was a significant effect for trials ($F(2, 53) = 7.34, p = 0.002$), meaning that participants were faster over time. On average, each subsequent attempt reduced the time by roughly a minute, so the average completion time for Trial 1 was 6.44 min, 5.12 min for Trial 2, and then 4.29 min for Trial 3. There was no significant correlation between training times and completion times for people who were successful or associations for those who crashed.

In the survey, participants were asked to label a grid-based map of the experiment setup for any areas that they found to be particularly challenging. This feedback was plotted as a heatmap on the map of the experiment environment (Fig. 6). Participants labeled the entrances of the shaft

Table 4: Success percentages and completion time means and standard deviations (minutes) for all three trials.

	Group 1 MC-SC/SC-MC		Group 2 SC		Group 3 SC+		Overall	
	%	Times	%	Times	%	Times	%	Times
Trial 1	0.29	6.22±1.02	0.36	6.21±0.93	0.50	6.74±1.31	0.38	6.44±1.09
Trial 2	0.64	5.02±1.76	0.71	5.30±1.47	0.57	5.06±2.14	0.64	5.14±1.72
Trial 3	0.63	3.64±1.74	0.62	4.77±2.40	0.67	4.21±1.12	0.64	4.29±1.80

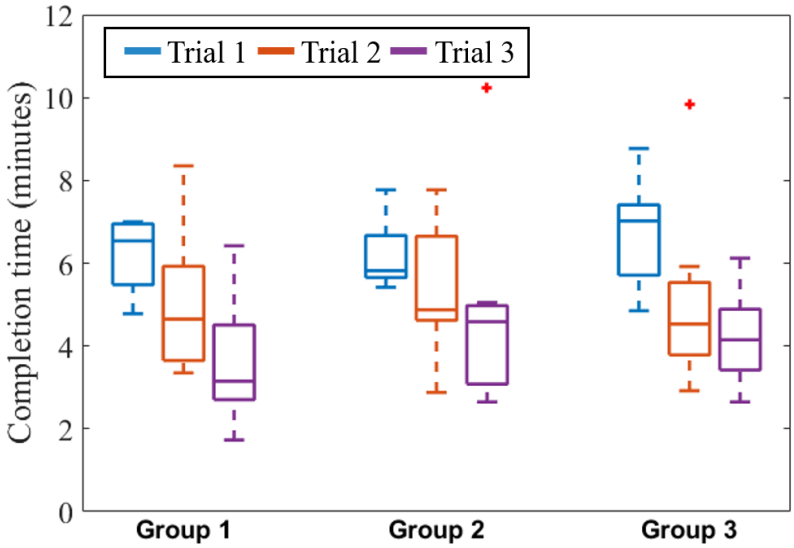


Figure 5: Mission completion times of the three trials.

Table 5: Details on the crashes for all three trials

	Trial 1	Trial 2	Trial 3
Overall crash rate for all groups	0.47	0.31	0.30
Crashes rate in the 1st half trip	0.71	0.92	0.80
Crashes rate in the 2nd half trip	0.29	0.08	0.20
Shaft-related crashes rate	0.67	0.42	0.50
Crashes rate into obstacles	0.19	0	0
Crashes rate into wall	0.14	0.58	0.50

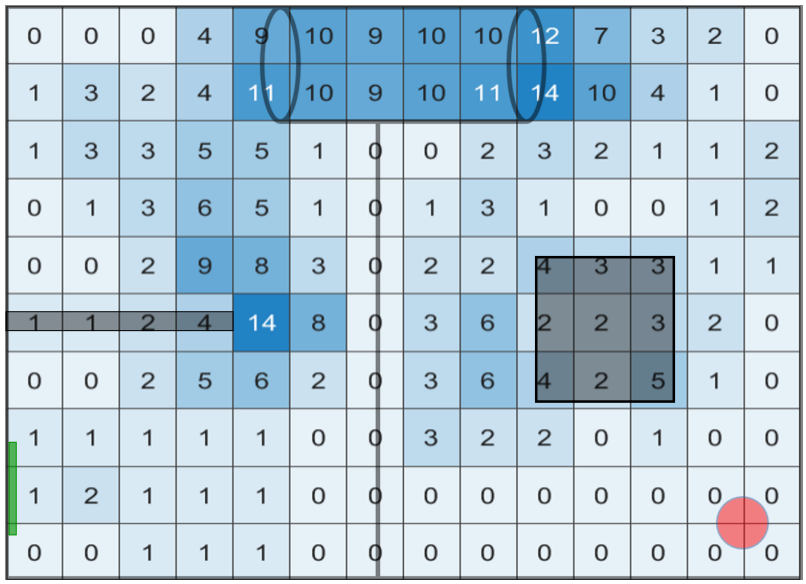


Figure 6: Heat map showing challenging regions. The number in a grid indicates how many times this grid was marked as challenging. The red dot is the home position where the UAV took off and landed.

and corner of the center wall as challenging regions. Interestingly, these two places were ranked almost equal in difficulty, although there were far fewer crashes on the center wall (1) as compared to 12 crashes at the shaft entrance 6. This suggests that people were not able to accurately assess areas of physical risk, likely due to the limited field of view or inadequacy of training.

Going through the shaft was the most difficult maneuver. Ideally, participants would make sure the path was clear with a brief look using the nudge control mode, then set a waypoint at the shaft entrance and fly to this point in waypoint control mode. Once at the shaft, participants should have switched to nudge control for fine-tuned control. However, many users tried to navigate through the narrow shaft by setting many repeated waypoints. This caused delays, added workload, and excessive use of power, which could lead to battery failure.

Nudge control could also be used inappropriately. It was designed to be used for close-in inspection and briefly checking to make sure paths were clear of obstructions. However, while observing people negotiate the test environment, we noticed that some people would almost exclusively use nudge control to not just inspect, but also navigate (contrary to instruction). While only 10% of people who were successful exhibited this tendency, 29% of people who failed tended to use nudge control inappropriately. There was no statistical difference in which groups these people were in.

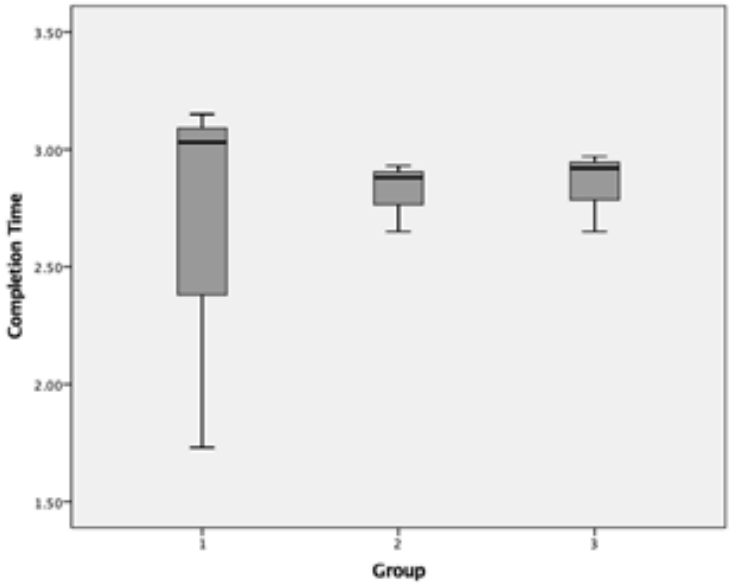


Figure 7: Box plots of top three performers in Groups 1, 2 and 3 that show average completion times and the interquartile ranges.

Given that the previous analyses focused not just on the average performances of those who succeeded, but also on the people who failed (the worst performers), we also examined the top performers. Fig. 7 plots the top three performances from Groups 1, 2, and 3 across all trials. These top performance completion times were more than 1 standard deviation faster than the overall mean time of 5.19 mins. The range of the top three times in Group 1 was 1.42 mins, while it was 0.28 mins and 0.32 mins, respectively for Groups 2 and 3. The variance of Group 1 was larger than Groups 2 and 3.

3.5. The Role of Overconfidence

In the post-flight survey, all performers were asked to rate their own performance after the first trial on a five-point Likert scale (very poor, poor, satisfactory, good, excellent). To examine the possible role of overconfidence, We then compared these assessments with participants' actual performances. Each participant's performance was classified into one of three categories: 1) Good if the flight was completed, 2) Satisfactory if the participant successfully read out the information on the target panel but crashed upon return; or 3) Poor if there was a crash or battery depletion prior

Table 6: Summary of participants' actual performance in each performance category in the first trial.

	Good	Satisfactory	Poor
Group 1	4	3	7
Group 2	5	5	4
Group 3	7	2	5

to reading the wall display. Table 6 summarizes how many participants' had good, satisfactory, or poor performance in Trial 1.

Then we compared these actual performance to their subjective assessments on a scale of -2 to 2, called the "Confidence Delta". On this new comparative scale, if participants' self-assessments matched their actual performances, they were given a 0, which we term "calibrated". If people reported that they performed excellent, but crashed, they were rated as very overconfident (a 2 since their perceptions were 2 places away from their actual performances). Similarly, they could report feeling their performances were poor, but actually completed the course, which would give them a rating of -2 (2 places below their actual performances) or very underconfident. Slightly overconfident = 1 and slightly underconfident = -1. Figure 8 illustrates this comparison that occurred after the first trial. It is the histogram of participants' confidence deltas in each group. Among each group, participants' confidence was also separately summarized based on their success and failure.

Overall for those who failed, 65% were slightly to very overconfident, compared to no cases of any overconfidence for successful trials, as seen in Fig. 8. Only one failure (4%) was slightly underconfident, but 31% of successes were slightly to very underconfident. A logistic regression model with the different training groups, confidence deltas, nudge preferences, ages, and hands-on training times used to predict success or failure yielded only the confidence delta as a significant variable ($Beta = -2.69, p = 0.008$). This model predicts that if overconfidence increased by one unit, participants were 15 times more likely to crash.

One participant from Group 1b reported significantly more UAV experience than any other participant (above 50 hrs UAV flying experience). This participant used 46.2 minutes and 23.5 minutes to complete the overall training and hands-on flight training, which were lower than the average times spent by Group 1 overall. However, this participant failed in Trials 1 and 2 and did

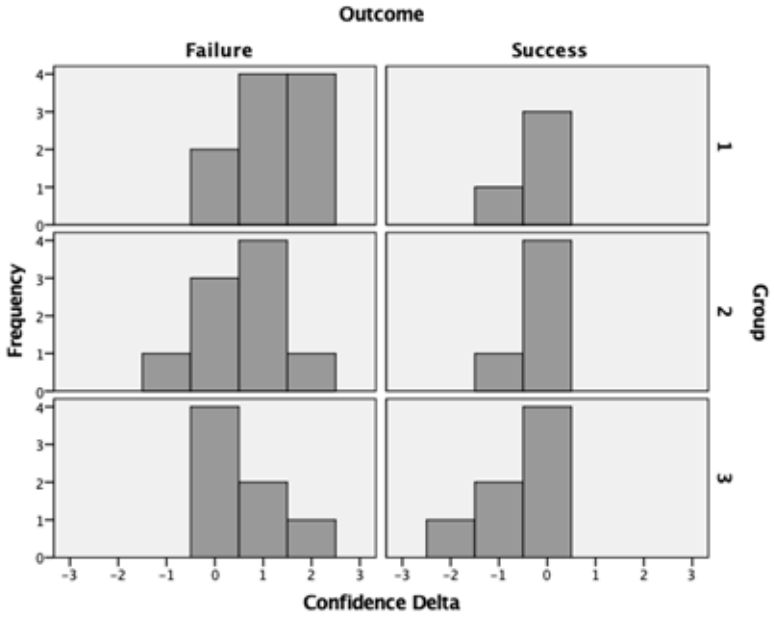


Figure 8: Confidence deltas between self-assessed performance and actual performance for successes and failures, by group.

not continue on Trial 3. Thus, the person with the declared most experience failed to complete a single test session, even though after the testing, he rated his performance as good. He was ranked as overconfident (2).

We investigated whether there was a control preference for those people who were overconfident, and there was a statistically significant positive moderate Spearman rho correlation ($\rho = .419, p = 0.001$) for people who preferred to use nudge control when navigating. Similarly, there was a weaker negative Spearman rho correlation ($\rho = -.289, p = 0.03$) for trust in waypoint control. This means that people who inappropriately favored nudge control in navigating also tended to be overconfident.

Understanding that overconfidence was a significant factor in this experiment, Fig. 9 illustrates what the overall crash rates across all three experiments were for the three groups, as well as the percentage of overconfident people in each group. Figure 9 clearly shows that Group 1 was handicapped by a large percentage of overconfident people, while Group 3 had the least. This helps to explain, in part, why Group 3 had the most people in the group of performers that successfully completed all three trials.

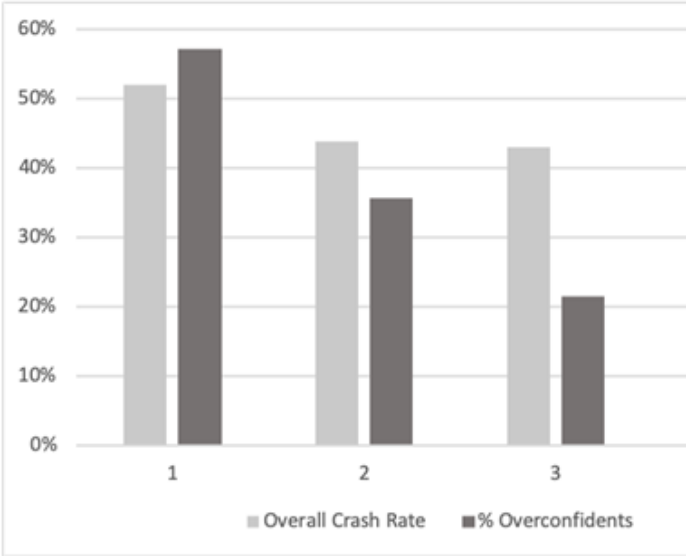


Figure 9: Crash rates and percentages of overconfident operators per group across the three testing trials.

4. Discussion

4.1. Training, Trust and Overconfidence

In the original experiment plan, Groups 1 and 3 had a time allotment of 120 minutes, and Group 2 had 82 minutes for training. However, participants of the three groups elected to use much less time to complete the training than allotted, on average 43.2%, 33.0%, and 28.7%, respectively, and only used 37%, 21.2%, and 22.8%, respectively of their allotted flight time. The original times for training were estimated using pilot studies, and a previous experiment [32]. In the self-evaluation of the quality of the training, thirty-nine participants (93%) believed the training was sufficient to prepare them for the test scenarios. Especially curious is Group 3 who had the option to have an extra 25 minutes of flight time, but chose not to take advantage of this extra time, and went even faster through the training. Such results raise questions surrounding self-paced training programs, which are the default for most recreational and commercial UAV training programs. In this study, participants did the bare minimum of training, which resulted in high crash rates in the initial test session (62%) and only dropped to 36% by the last session.

Those people who received manual control training (Group 1) received more training by default and reported their workload to be statistically lower than those without this training. This is

an especially important finding since a major theoretical benefit of supervisory control is reduced workload. In this study, people with manual control training had to work harder in training, but this extra training led to perceived workload reduction while testing.

While there was no obvious link between training time and performance, there was a much clearer link to performance through the measure of overconfidence. For people who failed after the first trial, 65% were either slightly to very overconfident, while none of the participants who successfully completed the first test session or successfully completed all three test sessions were overconfident. People who were overconfident were 15 times more likely to crash, and only those people whose confidence was appropriately calibrated or under-confident finished all three trials.

People's overconfidence can also be linked to their inappropriate trust and overuse of nudge control, as evidenced by the correlation between the two. It is not clear whether nudge control as a mode of control engenders overconfidence or whether overconfident people may trust nudge control more than waypoint control, this should be examined in future efforts. However, overconfidence was not always a negative trait as some people who were initially overconfident became top performers in later trials, as seen in Fig. 7, albeit with much more variation. These results illustrate the two sides to the overconfidence coin. Overconfidence was a handicap to most operators; however, in a small number of trials for operators with manual control training, overconfidence may have partially contributed to their overall fastest course times.

While these results indicate that many novice UAV pilots have difficulty estimating their abilities, the one person who reported significant UAV experience was also very overconfident and unable to successfully complete any test. We leave it to future work to determine how much overconfidence leads to poor UAV-piloting performance and how to develop training programs to appropriately calibrate people's confidence with their abilities, and if such calibration can lead to sustained high performance.

The derived measure of overconfidence indicates that some people had too much trust in their own abilities, which raised their risk of poor performance. This measure of trust in oneself was also significantly correlated to misuse of nudge control, manifesting in people using the camera to navigate. Such actions were not trained as they are akin to moving through the world while looking through a soda straw. Such behaviors suggest that this group of people (29%) distrusted the underlying waypoint supervisory control autonomy.

Additional results that showed that distrust in the waypoint control mode was also associated

with failed outcomes. These results are counter to results from a previous similar study where people appeared put too much faith in waypoint control despite the fact that the training content was the same between the two experiments [32]. Taken together, these results illustrate that too much trust as well as too little trust can contribute to poor outcomes. Designing for appropriate trust [36] is a well-recognized critical attribute of systems with embedded automation, which is further evidenced by these two experiments. Further work is needed to determine how such a trust level could be achieved in remotely-operated first-person UAV systems.

4.2. The Impact of Different Levels of Autonomy in Training

Our primary research question was whether providing people with manual control training would enhance performance with supervisory control. While there was no clear statistical conclusion that could be made about performance across the three groups, only people from Groups 2 and 3 successfully completed all three trials. None of these participants were overconfident.

While the overconfidence issue is likely the major contributor to crashes across all groups, it does not shed light on whether any specific training and autonomy combination produced superior performance for those who did not crash. Group 3 was statistically faster than Group 2 in terms of those who completed all three trials, and they were also faster overall for those people who were not overconfident. However, Group 3 performers were not necessarily the fastest performers. An overconfident person from Group 1 eventually flew the course the fastest in 1.73 minutes, which is impressive, but as a whole, Group 1's times were not as consistently fast as compared to the fastest people in Groups 2 and 3 (Fig. 7), who completed the course within 19s of each other. None of the fast operators in Groups 2 and 3 were overconfident.

Taking the crash rates and completion times together, there is no obvious best approach to training operators to fly first-person perspective UAVs. Ultimately, these results suggest that, like most engineering design decisions, there is a tradeoff in performance criteria. Figure 10 illustrates the complex relationship suggested by the data from this experiment. For those people that did well in this experiment (Fig. 7), those who had just supervisory control training (Groups 2 and 3) performed very well and were very consistent in their performances. The range in completion times across all successful trials was 7.2 minutes, so the fact that these performers were within 17-19s of their respective means is striking.

Those top performers that had manual control training in addition to supervisory control train-

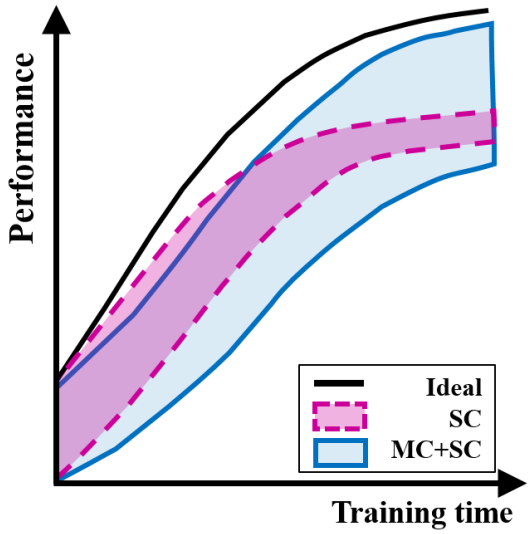


Figure 10: Hypothetical performance increasing patterns with supervisory control training (SC) only or combined training of manual control and supervisory control (MC+SC).

ing performed statistically no different than their counterparts with just supervisory training, however, their variability was much greater. The person with the fastest overall completion time had manual control training; however, collectively, this group was not as consistent as their range of scores varied by 1.42 minutes. The time for the second fastest person in Group 1 was 75% slower, a substantial drop. Moreover, while a minute and a half might not seem like a lot, in the use of small UAVs to perform visual inspection tasks, 1.42 minutes could mean the difference between success or failure in a mission.

One other important variable not explicitly captured in Fig. 10 is the cost of training. People in Group 1 spent, on average, 52 minutes training, while people in Groups 2 and 3 spent 31 minutes, so those with manual control training "cost" 68% more than just supervisory control training. Indeed, in one Air Force study that looked at the cost differences in having UAV operators learn to manually fly first before then learning to operate UAVs, as opposed to just learning to operator a UAV (similar to our MC+SC or SC hypothesis), they found that while manual control training overall made operators slightly better, they also more than tripled the cost [31].

So, one remaining question is whether it is worth the cost to develop extended training programs that teach manual control skills as well as supervisory control skills, if the overall net gain of doing

so may lead to excellent performance, but also potentially lead to a wider range of performances? Having a cheaper training program that produces consistent, albeit somewhat slower operators may be preferable to those training programs that take many more resources, with only occasionally significantly faster outcomes. As illustrated in Fig. 10, the ideal case is one where people learn quickly and achieve high and predictable performance. However, as seen by the strong negative correlation between more training and low workload, there is a benefit to operators such that some level of training reduces workload and stress. We leave it to future work to determine how to best achieve this, as well as how to identify and mitigate potential problems with overconfidence early in the training process.

4.3. Limitations

In this study, we recruited 42 subjects, which is large for this kind of study but still relatively small in terms of the overall population. This sample included novices primarily because the goal of the research was to examine training effects of autonomy on people with little to no experience. Participants self-selected based on interest, but in the future, the selection of participants based on trust in technology, self-confidence, varying degrees of exposure to drones, and video gaming experience could yield different results. Indeed, other research has shown that video game experience can both help [41] and detract [42] from UAV operator performance, so more work is needed in this area.

5. Conclusion

In this effort, we investigated whether three different UAV training programs with varying levels of autonomy would influence operators' performances and perceptions. We examined whether training for both manual and supervisory control was a better approach to training than just using supervisory control training. We also examined whether time in training was a significant factor. One of the surprising findings was that when given self-paced training, most participants quickly went through the training and did not utilize all the hands-on training time made available to them. Given that self-paced training programs are standard in recreational and commercial UAV certification, this study raises questions about the efficacy of such programs, and ultimately the safety of the operators.

Another important result from this experiment is the strong link between overconfidence and failure. Operators who were overconfident were 15 times more likely to crash, and there was no apparent connection between the degree of autonomy in training and overconfidence. However, there was a connection between overconfidence and the desire to constantly control the aircraft, suggesting distrust on the part of the overconfident operators. While a person's overconfidence is predominantly inherent, shaped by experiences with past technologies, it may be possible that training and onboard autonomy also influence confidence in this task. More research is needed to determine cause and effect as well as possible screening and mitigation strategies. This is especially important since commercial UAV licensing is expected to significantly increase in the next few years.

For those people who succeeded in their missions, the results were mixed. The overall top performer was a person who received manual control training before supervisory training. However, when comparing the top performers in this group with those top performers in the supervisory control groups, the results showed those people with just supervisory control training performed very consistently and quickly, although not as fast as the best performer. Ultimately, the sponsoring agency will need to decide the cost-benefit threshold, but these results suggest that when weighing such factors, supervisory control alone may provide sufficient skills for first-person UAV inspection tasks, assuming overconfidence can be mitigated.

6. Acknowledgments

This research was sponsored by the Office of Naval Research under the Science of Autonomy program. We also thank David Benac, Eddy Lin, Kausthub Ramachandran Kunisi Sekar, Rebecca Schmitt, Suman Bajgain, and Yao Yuan for their contributions to the experiments.

References

[1] H. Shakhatreh, A. H. Sawalmeh, A. Al-Fuqaha, Z. Dou, E. Almaita, I. Khalil, N. S. Othman, A. Khreishah, M. Guizani, Unmanned aerial vehicles (UAVs): A survey on civil applications and key research challenges, IEEE Access 7 (2019) 48572–48634.

[2] S. G. Gupta, D. Ghonge, P. M. Jawandhiya, et al., Review of unmanned aircraft system (UAV), International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 2.

[3] C. E. Nehme, J. W. Crandall, M. Cummings, An operator function taxonomy for unmanned aerial vehicle missions, in: 12th International Command and Control Research and Technology Symposium, 2007, pp. 19–21.

[4] M. Silvagni, A. Tonoli, E. Zenerino, M. Chiaberge, Multipurpose UAV for search and rescue operations in mountain avalanche events, Geomatics, Natural Hazards and Risk 8 (1) (2017) 18–33.

[5] P. Doherty, P. Rudol, A UAV search and rescue scenario with human body detection and geolocalization, in: Australasian Joint Conference on Artificial Intelligence, Springer, 2007, pp. 1–13.

[6] M. B. Bejiga, A. Zeggada, A. Nouffidj, F. Melgani, A convolutional neural network approach for assisting avalanche search and rescue operations with UAV imagery, Remote Sensing 9 (2) (2017) 100.

[7] T. Tomic, K. Schmid, P. Lutz, A. Domel, M. Kassecker, E. Mair, I. L. Grixa, F. Ruess, M. Suppa, D. Burschka, Toward a fully autonomous UAV: Research platform for indoor and outdoor urban search and rescue, IEEE robotics & automation magazine 19 (3) (2012) 46–56.

[8] K. Whitehead, C. H. Hugenholtz, Remote sensing of the environment with small unmanned aircraft systems (UASs), part 1: A review of progress and challenges, Journal of Unmanned Vehicle Systems 2 (3) (2014) 69–85.

[9] K. Whitehead, B. Moorman, C. Hugenholtz, Low-cost, on-demand aerial photogrammetry for glaciological measurement., Cryosphere Discussions 7 (3).

[10] W. Immerzeel, P. Kraaijenbrink, J. Shea, A. Shrestha, F. Pellicciotti, M. Bierkens, S. De Jong, High-resolution monitoring of himalayan glacier dynamics using unmanned aerial vehicles, Remote Sensing of Environment 150 (2014) 93–103.

[11] C. Deng, S. Wang, Z. Huang, Z. Tan, J. Liu, Unmanned aerial vehicles for power line inspection: A cooperative way in platforms and communications, Journal of Communications 9 (9) (2014) 687–692.

[12] Z. Li, Y. Liu, R. Walker, R. Hayward, J. Zhang, Towards automatic power line detection for a UAV surveillance system using pulse coupled neural filter and an improved hough transform, Machine Vision and Applications 21 (5) (2010) 677–686.

[13] V. Gonzalez-Dugo, P. Zarco-Tejada, E. Nicolás, P. A. Nortes, J. Alarcón, D. S. Intrigliolo, E. Fereres, Using high resolution UAV thermal imagery to assess the variability in the water status of five fruit tree species within a commercial orchard, Precision Agriculture 14 (6) (2013) 660–678.

[14] J. Primicerio, S. F. Di Gennaro, E. Fiorillo, L. Genesio, E. Lugato, A. Matese, F. P. Vaccari, A flexible unmanned aerial vehicle for precision agriculture, Precision Agriculture 13 (4) (2012) 517–523.

[15] D. Jenkins, B. Vasigh, The economic impact of unmanned aircraft systems integration in the United States, Association for Unmanned Vehicle Systems International (AUVSI), 2013.

[16] Federal Aviation Administration, FAA aerospace forecast fiscal years 2019-2039, Last visited: July 14, 2019 (2018).
URL https://www.faa.gov/data_research/aviation/aerospace_forecasts/media/FY2019-39_FAA_Aerospace_Forecast.pdf

[17] V. Rodríguez-Fernández, H. D. Menéndez, D. Camacho, Automatic profile generation for UAV operators using a simulation-based training environment, Progress in Artificial Intelligence 5 (1) (2016) 37–46.

[18] Y. Khosiawan, I. Nielsen, A system of UAV application in indoor environment, Production & Manufacturing Research 4 (1) (2016) 2–22.

[19] R. Joslin, Synthesis of unmanned aircraft systems safety reports, Journal of Aviation Technology and Engineering 5 (1) (2015) 2.

[20] A. V. Fernando, Survey of sUAS unintended flight termination as depicted in internet video, Journal of Unmanned Vehicle Systems 5 (3) (2017) 109–114. doi:10.1139/juvs-2017-0007.

[21] G. Polek, Drones offen lack FAA approval and fly too high, says report, Last visited: June 17th, 2020 (2020).

URL https://www.ainonline.com/aviation-news/business-aviation/2020-05-20/drones-often-lack-faa-approval-and-fly-too-high-says-report/.

[22] R. Parasuraman, K. A. Cosenzo, E. De Visser, Adaptive automation for human supervision of multiple uninhabited vehicles: Effects on change detection, situation awareness, and mental workload, Military Psychology 21 (2) (2009) 270–297.

[23] D. B. Kaber, M. R. Endsley, The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task, Theoretical Issues in Ergonomics Science 5 (2) (2004) 113–153.

[24] T. R. Carretta, Predictive validity of pilot selection instruments for remotely piloted aircraft training outcome, Aviation, Space, and Environmental Medicine 84 (1) (2013) 47–53.

[25] B. T. Schreiber, D. R. Lyon, E. L. Martin, H. A. Confer, Impact of prior flight experience on learning predator UAV operator skills, Tech. rep., Air Force Research Lab Mesa AZ Human Effectiveness Directorate (2002).

[26] S. Biggerstaff, D. Blower, C. Portman, A. Chapman, The development and initial validation of the unmanned aerial vehicle (UAV) external pilot selection system., Tech. rep., Naval Aerospace Medical Research Lab Pensacola FL (1998).

[27] A. Mairaj, A. I. Baba, A. Y. Javaid, Application specific drone simulators: Recent advances and challenges, Simulation Modelling Practice and Theory.

[28] J. T. Hing, P. Y. Oh, Development of an unmanned aerial vehicle piloting system with integrated motion cueing for training and pilot evaluation, Journal of Intelligent and Robotic Systems 54 (1-3) (2009) 3–19.

[29] J. T. Hing, K. W. Sevcik, P. Y. Oh, Improving unmanned aerial vehicle pilot training and operation for flying in cluttered environments, in: 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, 2009, pp. 5641–5646.

[30] J. Menda, J. T. Hing, H. Ayaz, P. A. Shewokis, K. Izzetoglu, B. Onaral, P. Oh, Optical brain imaging to enhance UAV operator training, evaluation, and interface development, Journal of Intelligent & Robotic Systems 61 (1-4) (2011) 423–443.

[31] G. Zacharias, M. Maybury, N. Cooke, M. Cummings, M. Draper, K. Fall, P. Worch, Operating next-generation remotely piloted aircraft for irregular warfare, Tech. rep., Technical Report. United States Air Force Scientific Advisory Board (2011).

[32] J. Zhou, H. Zhu, M. Kim, M. L. Cummings, The impact of different levels of autonomy and training on operators' drone control strategies, ACM Transactions on Human-Robot Interaction (THRI) 8 (4) (2019) 1–15.

[33] J. D. Lee, N. Moray, Trust, self-confidence, and operators' adaptation to automation, International journal of human-computer studies 40 (1) (1994) 153–184.

[34] J. D. Lee, Review of a pivotal human factors article:"Humans and automation: use, misuse, disuse, abuse", Human Factors 50 (3) (2008) 404–410.

[35] R. Parasuraman, V. Riley, Humans and automation: Use, misuse, disuse, abuse, Human factors 39 (2) (1997) 230–253.

[36] J. D. Lee, K. See, Trust in technology: Designing for appropriate reliance, Human Factors 46 (1) (2004) 50–80.

[37] J. Y. Chen, M. J. Barnes, Human–agent teaming for multirobot control: A review of human factors issues, IEEE Transactions on Human-Machine Systems 44 (1) (2014) 13–29.

[38] S. Losey, Investigators: Pilot 'overconfidence' led to reaper crash in afghanistan, Last visited: May 29, 2019 (2017).
URL https://www.airforcetimes.com/news/your-air-force/2017/04/14/investigators-pilot-overconfidence-led-to-reaper-crash-in-afghanistan/

[39] D. Pitman, M. L. Cummings, Collaborative exploration with a micro aerial vehicle: A novel interaction method for controlling a mav with a hand-held device, Advances in Human-Computer Interaction 2012.

[40] J. M. Twenge, W. K. Campbell, B. Gentile, Generational increases in agentic self-evaluations among american college students, 1966–2009, Self and Identity 11 (4) (2012) 409–427.

[41] J. M. Wheatcroft, M. Jump, A. L. Breckell, J. Adams-White, Unmanned aerial systems (UAS) operators' accuracy and confidence of decisions: Professional pilots or video game players?, Cogent Psychology 4 (1) (2017) 1327628.

[42] M. L. Cummings, C. Mastracchio, K. M. Thornburg, A. Mkrtchyan, Boredom and distraction in multiple unmanned vehicle supervisory control, Interacting with computers 25 (1) (2013) 34–47.