Exploring 3D Gesture Metaphors for Interaction with Unmanned Aerial Vehicles

Kevin P. Pfeil
University of Central Florida
Orlando, FL
kevin.pfeil@knights.ucf.edu

Seng Lee Koh
University of Central Florida
Orlando, FL
ksenglee@knights.ucf.edu

Joseph J. LaViola Jr.
University of Central Florida
Orlando, FL
jjl@eecs.ucf.edu

ABSTRACT

We present a study exploring upper body 3D spatial interaction metaphors for control and communication with Unmanned Aerial Vehicles (UAV) such as the Parrot AR Drone. We discuss the design and implementation of five interaction techniques using the Microsoft Kinect, based on metaphors inspired by UAVs, to support a variety of flying operations a UAV can perform. Techniques include a first-person interaction metaphor where a user takes a pose like a winged aircraft, a game controller metaphor, where a user's hands mimic the control movements of console joysticks, "proxy" manipulation, where the user imagines manipulating the UAV as if it were in their grasp, and a pointing metaphor in which the user assumes the identity of a monarch and commands the UAV as such. We examine qualitative metrics such as perceived intuition, usability and satisfaction, among others. Our results indicate that novice users appreciate certain 3D spatial techniques over the smartphone application bundled with the AR Drone. We also discuss the trade-offs in the technique design metrics based on results from our study.

Author Keywords

3D Interaction; User Studies; Robots

ACM Classification Keywords

I.2.9 Robotics: Operator Interfaces; H.5.2 User Interfaces: Interaction Styles

General Terms

Design, Experimentation

INTRODUCTION

Human Robot interface design is becoming an increasingly important topic as the robot industry matures. One approach to these interfaces is through the use of upper body 3D interaction techniques. There is opportunity for a more natural and intuitive user interface involving such techniques, or as part of an overall multi-modal interface. An interaction technique based on 3D gestures or poses, especially techniques that do not require touch or wearable devices, is more desirable due to its ease of reconfigurability and programmability [6].

Figure 1. Giving a 3D interaction command to the UAV

One approach in 3D interaction design that facilitates understanding of the techniques is the use of metaphors, which are prevalent in many commercial games using motion control, such as WarioWare: Smooth Moves[1] and Dance Central 2[2]. These metaphors could be used outside of entertainment, such as in Human Robot Interaction (HRI). However, we have not found much work that addresses the qualitative and quantitative attributes of 3D interaction, specifically 3D gestures, within HRI. A compelling reason for this lack of work could be that incorporating metaphors into 3D interaction design, or the design process itself, is usually technology-driven as opposed to user-driven. Most of the user studies we found were also conducted in software-based simulations, such as video games, despite markedly dissimilar user experiences between live, virtual and constructive test-beds [1]. With the arrival of low-cost robotics such as the Parrot AR Drone and Roomba, we attempt not only to address usability with 3D gesture interfaces, but also explore the configuration and performance of what we perceive to be commonly understood metaphors applied to body gestures that can be implemented for both remote and proximate interaction with robots.

The type of 3D interaction on which we focus upon in this paper is direct and explicit teleoperations of robots with ap-

---

[1]http://www.nintendo.com/sites/software_warioware.jsp
[2]http://www.dancecentral.com/

plications to assistive robotics, or in military domains such as reconnaissance or search and rescue. We chose UAVs, given their extensive use in the aforementioned application domains, and their availability for domestic purposes mostly as entertainment apparatus. Taking inspiration from common gestures and physical control devices, we implemented a set of five upper-body gesture techniques to fly the UAV in all degrees of freedom using the Microsoft Kinect. With the release of the Microsoft Kinect 1.5 SDK onwards, both standing and seated skeletal tracking are supported even at Near Depth Range; this feature allows us to explore metaphors for gesture sets that perform similar tasks for a wider context of user profiles.

RELATED WORK

There has been a significant amount of literature reported on the successful extraction of gestures and poses or their subsets, such as hand/finger gestures, to communicate and perform Human Robot Interaction (HRI) tasks via 3D interaction techniques [2] [3] [11].

Guo et. al. explored interaction with robots using different tangible devices such as the multiple Wii-motes and the keyboard [4]. Although they did use motion control versus traditional input devices, they still relied on usage of additional hand-held technology to assign commands to their robot. We explore interaction with robots without any grasping or wearing any tracking devices.

Lichtenstern et. al. [7] explored commanding multiple robots using 3D selection techniques that do not require tangible devices. However, they did not explore different methods of 3D spatial interaction and solely reported their facilitation for selection and coordination of movements of multiple-robot teams.

Sugiura et. al. [10] implemented a multitouch interface with a mobile device, such as a smartphone, to operate a bipedal robot via finger swipes and taps. Finger contact on the display is performed with natural finger expressions or gestures that mimic the robot's bipedal actions such as walking, turning, kicking, and jumping. Our work focuses upon teleoperation of UAVs without physical contact with any tracking devices, and we also explore and analyse metaphors that may not be as straightforward as simple mimicry of the robot's locomotion via finger walking.

Ng et al. [8] explored a falconeering metaphor for interacting with an AR drone. However, they used a Wizard-of-Oz experiment set-up and with emphasis on studying human-robot social interaction inside collocated spaces. Our work is a full technical implementation using the Microsoft Kinect SDK and we do not differentiate interaction techniques for either collocated or remote teleoperations, with the gesture sets pertaining to each metaphor applicable to both sets of scenarios.

There have been many retrospective works reported on using body gestures to control robots, but they emphasize more on technical implementations for prototyping purposes [9] [10] [11]. Instead, we report the results of qualitative and quantitative metrics in a user study for the teleoperation of UAVs.

3D INTERACTION TECHNIQUES

Design Process

We aim at developing multiple 3D gestural interfaces based on metaphors that we regard as natural when applied to UAVs. By observing the physical nature and movement of the UAV, we were inspired to develop interaction techniques. We ultimately developed five interaction techniques to study. We regard these developed techniques as natural and hoped the participants also found these as natural.

Keeping in mind that users would be holding poses, we attempted to factor in parameters that would affect the overall usability of the technique. For instance, we believe that the neutral position, or the "do nothing" command, should be the least strenuous activity for the user. In other words, if a pose must be met by the user for the UAV to keep its course, the user should feel as comfortable as possible. Further, issuing commands to the UAV should not be difficult to do or require much energy; they should be easy for the user to perform and simple enough so that they become second nature. We wanted the reaction time of the UAV to be as fast as possible, so we did not use gesture recognition algorithms or complex gestures. Rather, we use specific heuristic values to allow for faster results.

After deciding on the interaction techniques based on the chosen metaphors, we developed the gestural commands to operate the UAV. The next sections describe the metaphors and the physical interactions that were used to control the UAV.

Selected Metaphors

Five techniques in all were created and assigned a moniker to assist the user in remembering the underlying motions to control the Drone:

- First Person

- Game Controller

- The Throne

- Proxy Manipulation

- Seated Proxy

We consider these five techniques to be built on easy metaphors to understand, and we attempt to generate command lists that follow the theme of the metaphor and are easy to use. See Figure 2 for examples of each technique.

First Person

This technique is based on the metaphor of the user assuming the pretense role of an aircraft. Children at play can sometimes be found mimicking an airplane, with their arms out to the side, "flying" around a playground. The technique was built to mirror this seemingly natural pose, where the user keeps the arms out to the side as if they were aircraft wings. To move the UAV to either side, the user would lean in the corresponding direction; to turn the Drone, the user would simply rotate the torso to the appropriate direction. To have the UAV climb or fall, the arms would respectively go up above the head or down below the waist. When the user leans forward or backward, the UAV would move in that direction.

We observe that it is possible for the user to give complex commands to the UAV using this method with low rates of accidental input. Due to this seemingly natural metaphor we expected performance to be positive. We also expect this technique to be favourable due to its simplicity.

Game Controller

This technique was developed using the metaphor where the user's arms assume the role of a typical set of control sticks found on a game controller. In many first person shooter games available for consoles, the left joystick traditionally controls translation of a character while the right joystick commands rotation. In this interface, the user's forearms operate in a similar fashion. When designing the interface, we originally intended for the hands to be in the air, where the elbows are bent at 90 degree angles. We quickly realized the amount of strain required to keep the arms in the air, so we ultimately rotated the entire interaction to counter this. As a result, the neutral position for the user is to hold the arms as if they were resting on a chair, but in a way that the shoulders are relaxed. When the left arm is angled right or left, the UAV would move in the corresponding direction. The right arm, when moved in the same way, would command the UAV to change its heading, respectively. To allow for vertical climb and fall, the right arm is raised or lowered, respectively, and to allow forward and backward translation, the left arm is moved in the same fashion.

Although this technique seems to involve a weaker metaphor, we wanted to explore this type of interaction where minimal movements of the user can still issue commands to the UAV. With these gestures, the user has more opportunity to assign complex commands to the UAV because both hands are independent of each other. Our expectation is that due to the minimal movement required to issue commands and the ease of combining commands, this technique would be favourable.

The Throne

We formed this interaction technique on the metaphor of the user assuming the role of a king or queen. The typical depiction of a monarch giving commands to a subject is always through the use of one hand, using minimal amounts of energy. For this technique, the user is seated in an armchair. The arm rests are used for the neutral position, meaning no command is given. Using one hand only, the Drone can be navigated up, down, left, right, and forward by simply pointing in the respective direction. The command to move back is to bring the hand towards the shoulder, as if to beckon the Drone backwards. Rotation for the Drone can be achieved by rotating the opposite hand to the left and right.

For this technique, the hand movements seem very simple. We expected this technique to be not only the most comfortable technique, due to the lack of strain on the user's legs, but the easiest one to use, because one hand can control the UAV almost entirely.

Proxy Manipulation

This technique was built on a metaphor that allows the user to visualize moving the Drone as though it was grasped in the user's hands. The hands are placed in a comfortable area directly in front of the user this is the neutral position. In order to manipulate the Drone, the user would move the hands forward, indicating the intent to push it in that direction; to bring it backward, the user would make a pulling gesture by bringing the hands near the shoulders. Turning the UAV in a direction involved moving one hand forward and the other hand backward, as though it was being steered. The user would lift the hands to raise the UAV, and conversely lower them to let the UAV descend. In order to strafe left or right, the user would position the hands one over the other, as if the UAV was being tilted in the corresponding way.

The concept of holding an imaginary UAV seems to be a very easy metaphor for a user to understand, and we expect highly favourable results from this technique. Because all commands are given in tandem using both hands, misplacing one hand typically renders no action by the UAV; we find this favourable because it helps to prevent accidental commands being given through errant gestures.

Seated Proxy Manipulation

This final technique was generated as an alternative for the original Proxy Manipulation method. The user takes a seated position, and most commands match the original s However, the left and right strafing is conducted in an different fashion. Opposed to applying imaginary tilt through the use of one hand over the other, the strafing is conducted by using both hands to pull the Drone to either side. This way, all commands are given to the drone by keeping the hands level on one plane.

We expected this technique to allow for a similar experience as the Proxy Manipulation, with the only exception in the comfort factor; it seems obvious that seated gestures would be preferred due to the lack of strain on the user's legs.

Parrot AR Drone Smartphone App

In addition to our interaction techniques, the user could also interact with the UAV using the native smartphone app developed by Parrot. We use an Android device containing an accelerometer for our study. To have the UAV rotate, the right thumb swipes on the screen and holds until the desired angle is achieved. Height can be adjusted by swiping up or down. To translate the UAV, the left thumb is held on the screen and the smartphone is physically rotated in the desired direction, affecting the accelerometer component that is used to generate commands.

USER STUDY

The goal of our study is to measure performance of the interaction techniques we developed, and evaluate parameters that may be important to the user when interacting with UAVs. The techniques are intended for rotorcraft, but similar variants can apply to fixed-wing UAVs. We aim to begin to answer the following questions:

- Are there specific interaction techniques that are more effective than others?

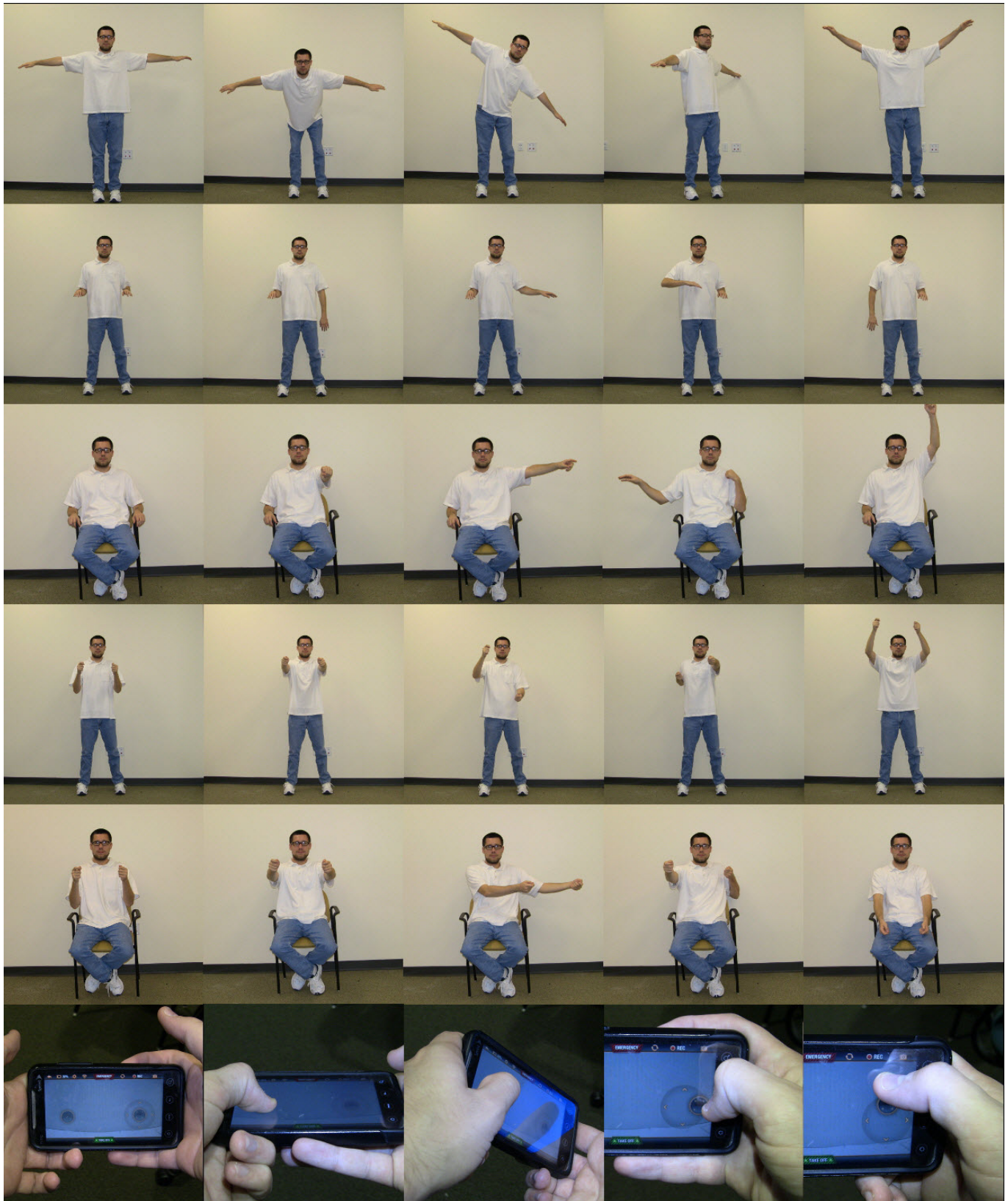- How are specific interaction techniques more ergonomically appreciated?

Figure 2. Examples of the various poses for each of the interaction techniques. From Top to Bottom: First Person, Game Controller, The Throne, Standing Proxy Manipulation, Seated Proxy Manipulation, Smart-phone App. Commands shown are examples of neutral positions, forward/backwards, strafe left/right, turn left/right, and up/down, in order per technique.

- Can 3D interaction outperform traditional control devices to control UAVs?

- Do more elaborate metaphors allow for easier understanding of the interaction?

We expect to find at least one technique that is not only more effective than others, but also more natural, among other factors. We also expect to find at least one technique that outperforms the default input method that acts as the control variable, in this case a smartphone application.

Subjects

14 students (10 male and 4 female) from the University of Central Florida were recruited to participate in the study. 11 are graduate students. The ages ranged from 20 to 37 and the median age is 28. Only 2 have ever interacted with the AR Drone prior to the user study, but half reported prior experience using remote controlled vehicles. 10 of the participants have used a Microsoft Kinect before.

Devices and Software

We selected the Parrot AR Drone 2.0 to serve as our test UAV. The AR Drone is a quadrotor vehicle equipped with two cameras: one facing forward and one facing downwards. It is bundled with a smartphone app that allows joystick-like control; our hand-held device was a HTC Evo 4G that has an accelerometer required to work the application. We used skeletal data extracted with Microsoft Kinect SDK to develop our interaction techniques. In order to push the commands to the AR Drone through the Kinect, we made use of an open source C-sharp project developed by Endres, Hobley, and Vinel[3], which allows PC communication with the Drone, and developed our interfaces in it. We ran this application on a computer running the 64-bit Windows 7 Operating System, with 8 GB RAM, a 2.3GHz Intel I7 processor, a 1920 x 1080 resolution on a 17" screen, with a 2GB NVIDIA GeForce GTX 660M graphics card. Because the software application did not display the camera feed for the AR Drone 2.0 at the time of the conducted study, we used FFPlay[4] to rapidly decode the image stream and display it on the laptop.

Test Space

To conduct our study, we arranged for a space of approximately 15m long, and 6m wide. No obstacles were within this safe area, but there were walls and objects outside of the border. The user was located approximately 4.5m from the shorter side, with the Kinect stationed in front of the user. Although this can create parallax error due to the length, we decided that it was beneficial for the user to see the entire space directly in front, requiring a smaller field of view. In each corner of the test space, way points were placed with a 1.2m by 1.2m area. The Drone's beginning and ending location was another way point located in the middle of the two long sides, but closer to the user, for an easier view of the Drone's take-off and land. Near the middle of the long side of the space, a visual target was placed outside of the boundaries,

[3]https://github.com/shtejv/ARDrone-Control-.NET
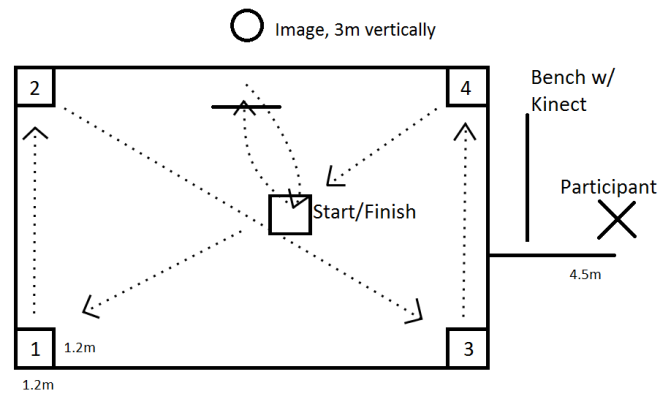[4]http://ffmpeg.org/index.html

Figure 3. Layout of the test space for the user study. The user was placed away from the area for a full view of the environment. The UAV was to navigate from the start to the numbered way points in order. The UAV was required to turn towards the user before moving from point 2 to point 3. After the Drone arrived back at the start, it was to look at the image outside of the test area before landing at the finish.

approximately 3m in height. This was used as part of the experimental tasks, which required the camera of the Drone to view the image.

Trial Design

Participants used all techniques in a random, unique order. Before performing a run, the user was given ample time to become familiar with the interaction technique. Up to 5 minutes of free usage was allowed before starting the timed trial, but the user could opt to begin before 5 minutes elapsed. During our initial pilot studies, expert users were able to complete the course with any technique in just over one minute. We feel that the 5 minutes of training time, as well as a 5 minute maximum to complete the course, is ample time for a successful run. During familiarization, all commands were tested by the user and inverted flying was suggested.

In order to evaluate user ability to navigate the AR Drone with each described technique, we involved the way points of the test area and required the user to fly through them at any height. Because parallax error affects the user's ability to perceive the Drone's location at further distance, we used a human line judge to inform the user of successful way point achievement. The order of the way points was constant between users. Starting with the landing pad and working forward to the left, the Drone would fly in a "figure 8" with no navigational restrictions, with the exception of one key constraint. Upon reaching the second way point and needing to fly towards the other side of the area, the Drone was required to turn so that it faced the user; this forced an understanding of the inverted commands, as left/right and forward/backward were now switched, from the user's point of view. To assist with this task, the camera was available at all times on the laptop screen or on the smartphone. Once the figure 8 was completed, the Drone flew towards an image set-up beyond the boundaries of the test area. It did not need to cross the borders of the space; rather, it needed only to look directly at the image. The line judge would assure proper positioning of the Drone, or the camera feed was checked to ensure the im-

In-Between Questionnaire
The interface to fly the Drone was comfortable to use.
The interface to fly the Drone was confusing to me.
I liked using the interface.
The gestures to fly the Drone felt natural to me.
It was fun to use the interface.
I felt frustrated using the interface.
It was easy to use the interface.
The Drone always moved in an expected way.

Table 1. Questions asked to the user for Rate and Rank.

Positive Feedback
I was impressed by how cool and novel it is.
The metaphor of holding the drone is easy to understand.
Fun and intuitive.
The controls were efficient.
Better control of the Drone.

Table 2. Positive comments captured by users for Standing Proxy.

age was being viewed. After completing all tasks, the Drone would proceed to land on the point where it began, with some buffer due to drift while landing. It was possible to fly the Drone too far out of bounds, but the user had a chance to recover without penalty. If a crash occurred due to losing control of the Drone or errantly clipping an object, the Drone was reset to the landing pad.

Quantitative Metrics
On successful runs, the total time from rotor start-up to rotor shut-down was recorded. Up to 3 attempts per interaction technique were allowed for a user. If the Drone crashed at any point, the Drone and timer were reset, and a new try was conducted. After any third failed run, the maximum time allowed, five minutes, was recorded for that run.

Qualitative Metrics
After using every interface, the participant was given a questionnaire to rate key qualities about the interface, according to their experience. Using a 7-point Likert scale, where 1 means strongly disagree and 7 means strongly agree, participants were asked the following questions listed in Table 1.

Each participant was also asked to fill a post-questionnaire at the end of the entire experiment. We asked the user to rank each technique based on the same criteria in Table 1 plus one additional question that asks the participants to rank their overall experience. This ranking ensures that there can be no ties in any category.

RESULTS AND DISCUSSION

Analysis of Trial Completion Times
Figure 4 illustrates the mean completion time of trials with each 3D interaction technique and the smartphone app. If a participant could not complete the course within 3 tries, we gave the cap of 5 minutes as a penalty for that technique, and all participants who finished the course did so under the 5 minute cap. Only 1 participant could not complete the course with any technique. Except for the Throne, each technique had either one or no participants who failed the course.

We used a 6-way repeated measures ANOVA analysis to test for significant differences in the mean completion times between all interaction techniques and smartphone app in our study. If there were any significant differences found between the groups, we use matched-pair t-tests to look for interesting differences between any 2 sets of interaction techniques in our post-hoc analysis. For instance when using the smartphone

app as the control, we perform a t-test comparison with each of the other techniques at $\alpha=.05$, resulting in a total of five comparisons. Type I errors inside the t-tests are controlled by using Holm's Sequential Bonferroni adjustment [5].

Significant differences were found between the interaction techniques and smartphone app in their trial completion times ($F_{5,13} = 4.201, p < 0.002$). However, using pairwise t-tests with the smartphone app as the control, we found no significant differences with the interaction techniques. Completion times from 'The Throne' technique was the cause of the significant differences between the groups, due to less than half of the participants not being able to complete the trial. We proceeded to do pairwise t-tests with the 'The Throne' technique as the control instead. Significant differences between The Throne and the other interaction techniques were found, except for the smartphone app (Seated Proxy t=3.164, p < 0.007; Standing Proxy t=3.037, p < 0.01; First Person t=2.796, p < 0.015; Game Controller t=2.607, p < 0.022). Figure 4 implies that the mean completion time for the smartphone app is comparable with The Throne technique.

Although users had ample time to become familiar with The Throne, we believe the gestural commands were not well-tuned. Additionally, we perceive difficulty when users attempt to recover from unexpected UAV behaviour. Confusion paired with very simplistic gestural commands yielded poor performance.

ANALYSIS OF QUALITATIVE DATA
We used a non-parametric Friedman test on our post questionnaire qualitative metrics to check for any significant differences in their medians. If any significant differences were found between the groups, we use Wilcoxon signed rank tests to look for differences between the interaction techniques with the smartphone. Type I errors are controlled by using Holm's Sequential Bonferroni adjustment [5]. Quite similar to the quantitative time completion analysis, we found no significant differences in any the qualitative metrics, except for "Fun" ($\chi^2 = 14.776, p < 0.01$) and "Likeability" ($\chi^2 = 12.342, p < 0.03$). The interaction technique that was significantly different than the others was the Standing Proxy, which benefited from a greater appreciation from the participants. Table 2 contains comments collected from the users after their experience.

Overall, user feedback indicates the Proxy techniques to be the best out of the 5 developed 3D interactions. The smartphone application was regarded as the best interface by 4 of 14 participants, whereas the Standing Proxy was ranked the highest by 6 of 14. The Throne technique was regarded as
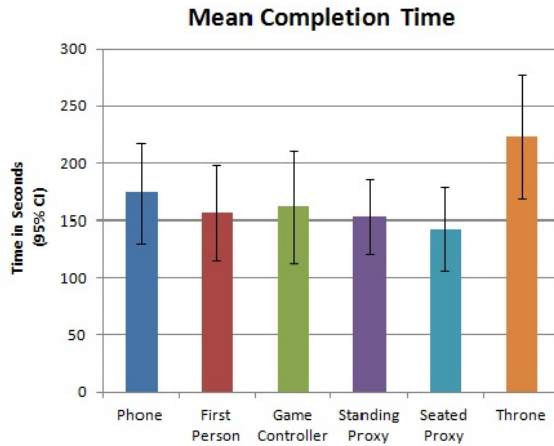
Figure 4. Mean completion time for each interaction technique. There were subtle differences in performance of all techniques, except for The Throne, which did not perform well.
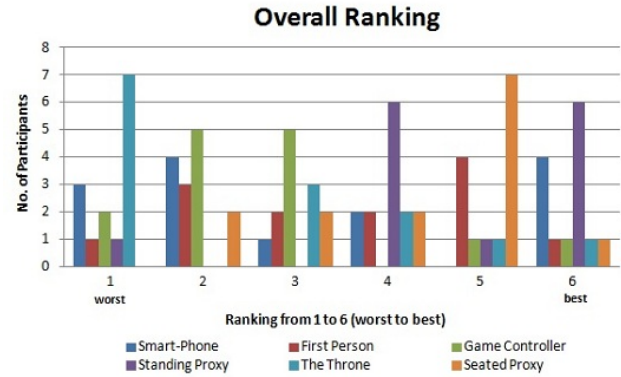

Figure 5. Histogram of user rankings for the interfaces overall. Users indicate that the Standing and Seated Proxy techniques are better to use than the others.

the worst technique by half of the participants. Figure 5 illustrates the rankings for each interaction technique.

The First Person technique felt the most natural, according to user feedback. According to some participants, because the underlying metaphor is very easy to understand, most of the commands did not need to be explained. Users also commented that they did not even think about what gesture to perform when they needed it; they felt it was second nature, as we expected. Figure 6 depicts the user rankings of perceived naturalness of each technique.

The Throne and the Game Controller techniques were both very confusing to the users as shown in Figure 7. As previously discussed, The Throne was the worst in terms of trial completion time, and it was also ranked very poorly overall by the users. That is due to the confusion the users perceived when attempting to recover from the UAV's incorrect movements. The Game Controller was also perceived as confusing, most likely due to the weaker metaphor that the technique was built on. If the neutral position was intended for the hands to be constantly in the air as originally designed, there may be opportunity for improved perception. However, the trade-off for this switch in stance would be any comfort currently achieved. Like The Throne, it is easy to give inadvertent commands due to the minimal hand movements needed to pilot the UAV.

The smartphone app was ranked the most comfortable more times than our 3D techniques due to the absolutely minimal movements needed to command the UAV. However, the Standing and Seated Proxy techniques were also regarded as comfortable, with the Standing Proxy being the better of the two; this is due to fewer movements needed to perform a strafe. We expected the Seated Proxy to be more comfortable due to less strain on the legs, but since interaction only occurred for up to 5 minutes, user fatigue may not have had time to reach a level that would have caused annoyance. The First Person technique was regarded as very uncomfortable, which is understandable, as the user needs to not only keep

the arms spread at all times, but also lean in different directions to command the UAV. Figure 8 depicts user rankings of perceived comfort levels of each technique.

Figure 9 details the overall rankings of the participant's disposition towards the factor of "Likeability." All but one of the users ranked the Standing Proxy technique positively. This user's comments include "turning was frustrating" but we observed the user performing the turning interaction task incorrectly. The other users had no problem with any of the commands and liked the interface overall. 9 of the 14 participants also ranked the Seated Proxy technique highly, as we expected due to the Standing and Seated versions being very similar. Comments from the participants indicate that they preferred the hand-over-hand method to strafe the UAV opposed to the horizontal movement for both hands. As illustrated in Figure 2, the strafing is conducted in a rather extreme manner. Transitioning between strafes left and right involves the user to move the arms from one side to the complete opposite side, whereas the Standing Proxy strafing only involves easier movements for the arms. This is an interesting find, as this helps show that even though the metaphor is easy to understand, the design of the actual commands still need to be considered.

Figure 10 depicts user frustration when using each interaction method. There was no clear advantage by any the methods that reduced frustration levels. Looking to user ratings for this metric, it seems that the median level of frustration, on the 7-Point Likert scale, is 2, with the exception of the Throne. This suggests that although there may have been some discomfort or confusion when using a technique, it was not enough to ultimately frustrate the user.

Half of the participants ranked the Standing Proxy technique as the most fun, as shown in Figure 11. Interestingly, the smartphone app was regarded as one of the least fun interaction techniques. We attribute this finding to the idea that exaggerated gestures may be regarded as more fun for a user, whereas a more simple mode of input does not necessarily bring enjoyment. We originally believed that the most fun interface would be that which yielded the best performance

in task time, but does not seem to be the case. In general, it seems that the use of motion control is more fun for a user than standard methods of input.

Figure 12 depicts ranking of user perception of the ease of use for each interaction technique. The Proxy techniques were generally regarded positively in this regard, with 9 of 14 users reporting the Seated Proxy positively, and 11 of 14 regarding the Standing Proxy technique positively. The smartphone app was also very easy to use, but some users did have trouble using it. We observed some users rotating the smartphone too far when trying to translate the UAV; in this way, the UAV would either not move in the desired direction, or it wouldn't move at all. User feedback indicates that the Proxy techniques were easy to understand and also easy to perform.

Figure 13 measures the ranking of the UAV's movements in an expected fashion while navigating it with each technique. Half of the users found the UAV moving expectedly when using the smartphone app. The Proxy techniques were again regarded positively, indicating that user gestures agreed with how the UAV was expected to move when navigating the course. The Throne technique did not allow for expected movements by many of 11 of 14 participants.

LESSONS LEARNED
Although we cannot directly generalize our results to other forms of robots, we expect that our methodology for developing the 3D interaction techniques used in this study can be applied to creation of future user interfaces. Metaphors are efficient facilitators for design teams to brainstorm potentially efficient techniques in 3D gestural user interfaces, as it helps both designers and users to form a natural cognitive mapping between perception of a system and its actual machine operations. By introducing metaphors to the designed gestural command sets, we can estimate the benefits and drawbacks for a technique in a known usage context. For instance, we did expect the First Person technique to perform well as it used a very simple metaphor to comprehend, but we also expected strain on the user's body. Our questionnaires confirmed our suspicion. Analysing the qualitative results of

each technique, our initial suspicions for each technique have been confirmed or denied.

First Person
We originally believed the First Person technique to be very favourable as an interface to command the UAV, but this was only true to an extent. Users did find this method to be natural, fun, and not confusing, but was ultimately regarded as a mediocre interface due to the amount of discomfort brought on by the excessive movements needed to give commands. We would not recommend this type of interface for prolonged usage, but perhaps for entertainment purposes. Paired with the AR Drone's front-facing camera, this would be a ideal interface to use when exploring an open world with a UAV.

Game Controller
Our initial reaction to the Game Controller technique was that it would allow an easier time to interface with the UAV in a more complex way. Since both arms can be used independent of each other and are used to give separate commands, more complex commands can be used to fly the UAV (such as turning while flying forward and climbing higher). By allowing the users to navigate the UAV in this way, we expected performance to increase. However, the participant feedback indicates that the interface is not as natural as the others, and it is also very confusing. Due to both arms being independent from each other, the user was able to give inadvertent commands in more than one way at a time. We observed users accidentally flying the UAV to the ground while attempting to fly in a straight line, which was caused by the user forgetting to maintain the neutral position. In prior pilot studies, experts were able to use this technique effectively, however; we would suggest that this technique be used only after ample training.

The Throne
From our results, we found the Throne technique to perform the worst; half of the users could not complete the course, and we attribute this poor result to the technique's requirement of very minimal movements from the user in order to operate the UAV. The Throne technique is essentially a pointing gesture,
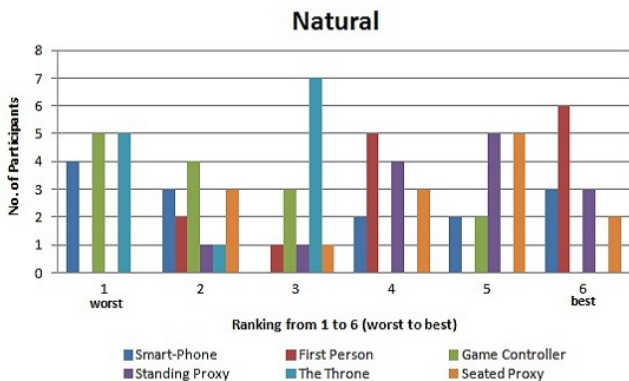

Figure 6. Histogram of user rankings of Naturalness. The First Person technique was perceived as the most natural to use by 6 of 14 participants.
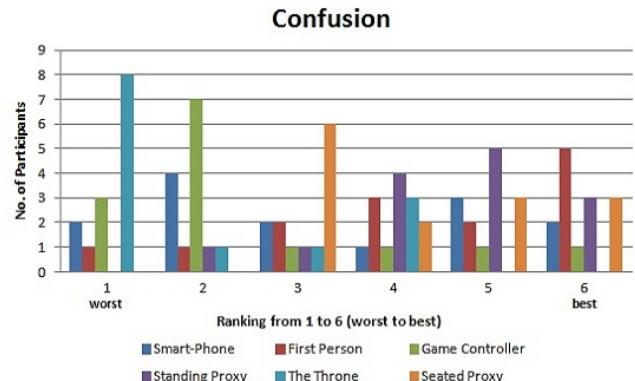

Figure 7. Histogram of user rankings of Confusion. The Throne was the most confusing for nearly all participants, and the First Person technique was regarded as least confusing.
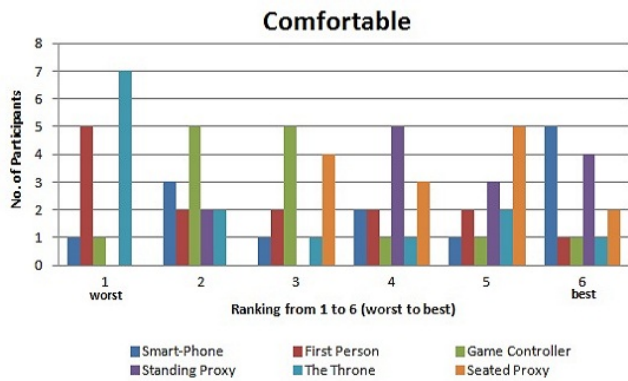
Figure 8. Histogram of user rankings of Comfort. The First Person technique was perceived as the most natural to use by 6 of 14 participants.
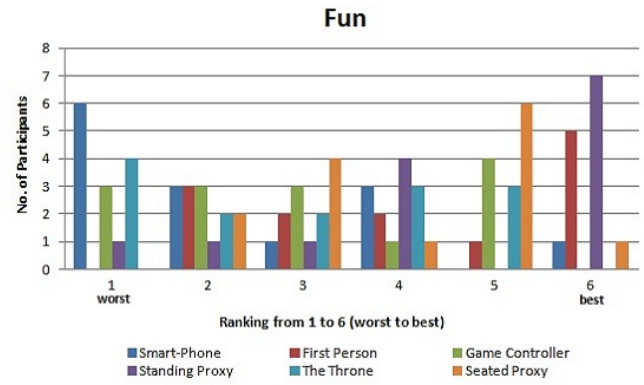

Figure 11. Histogram of user rankings of Fun. 7 of 14 participants indicate the Standing Proxy technique was the most fun, and 6 indicate the Smart-Phone App was the least fun.
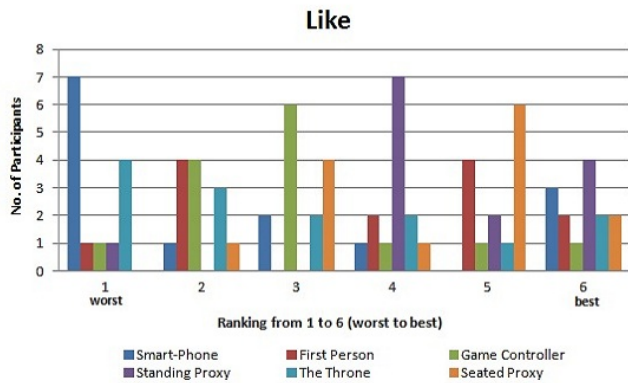

Figure 9. Histogram of user rankings of Likeability. All but one user ranked the Standing Proxy positively.
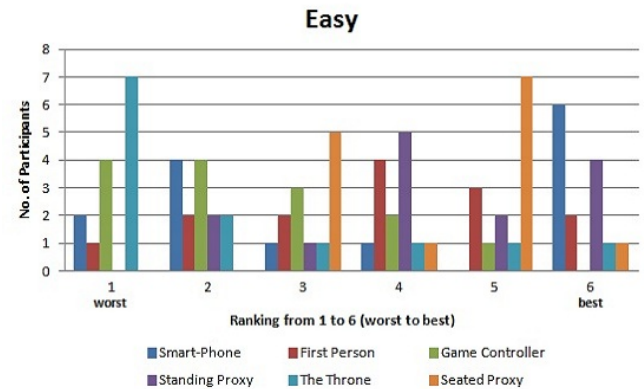

Figure 12. Histogram of user rankings of Easiness. The smartphone app was regarded as the easiest, and among the 3D interaction techniques, the Proxy techniques were regarded favourably.
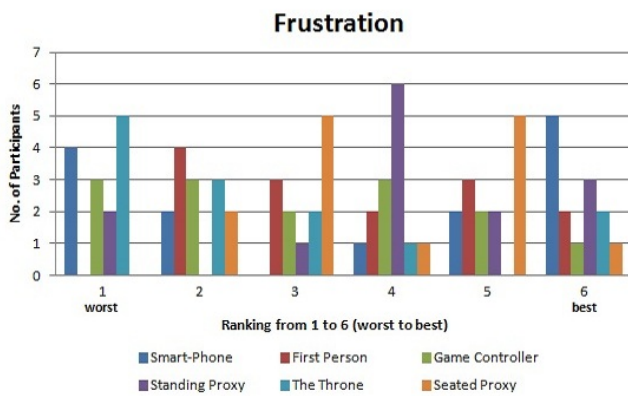

Figure 10. Histogram of user rankings of Frustration. The Seated Proxy technique was well regarded as a non-frustrating technique, and The Throne was generally most frustrating.
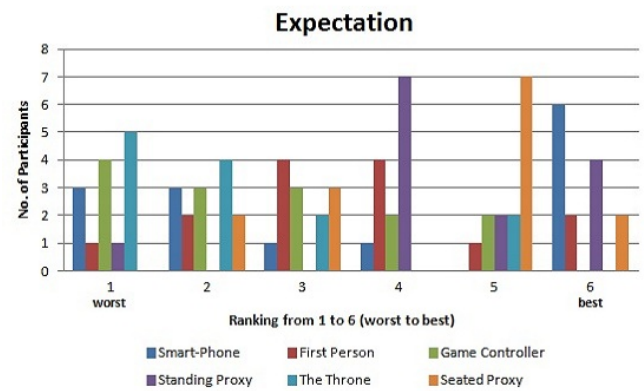

Figure 13. Histogram of user rankings of Expectation. The smartphone app was mixed with both positive and negative feedback; the proxy techniques were generally regarded as positive.

with some modification, and when a user becomes confused or the UAV moves towards an obstacle, incorrect commands can be very easily dictated. Similarly, inadvertent commands can be given by accidentally moving the hand in any direction. We do not expect this kind of technique to be useful in many domains due to the difficult nature of its use, and therefore would not recommend its use.

Standing and Seated Proxy

In contrast to The Throne and Game Controller, the Proxy techniques require both arms in tandem to perform any command. Although this somewhat requires more user effort, we find that accuracy is greatly improved, as inadvertent commands are less likely to occur. Incorrect commands are also reduced because the user is required to move the body more purposefully to the correct gesture for the desired command. Users regarded these techniques highly among all of the recorded factors, indicating that this style of interaction may be the best out of the ones we developed. We would therefore recommend use of these techniques over the others to navigate a UAV.

FUTURE WORK

We plan on further evaluating user preference over other techniques not reported here, and we also plan on attempting to find a technique that can potentially combine all of the positive attributes reported here by the test subjects. Although this study did compare 3D interaction to a control device, we plan on further comparing differences between 3D interaction techniques to other traditional input devices, such as the keyboard, joystick, and typical R/C controllers. Further, we want to explore the development of 3D interaction techniques for different robot platforms, including ground vehicles and humanoid robots, and attempt to find natural techniques to improve current HRI interaction standards. Lastly, we plan on applying our findings to multiple in-tandem robot teams instead of just one standalone agent.

CONCLUSIONS

We developed 3D interaction techniques using the Microsoft Kinect SDK to control the AR Drone 2.0. A user study was performed that evaluated the participants' disposition towards each interface on multiple levels. We find that users appreciate the designed techniques that are built on easy and understandable metaphors, which ultimately serves for better interaction. Our test subjects preferred our Proxy Manipulation techniques, regardless of the posture, over others that were still regarded as fun or perceived as natural. Due to varying results in the factors of comfort, likeability, naturalness, and overall perception, we conclude that there may be a correct usage for each of these techniques when applied in a proper domain; for instance, the Proxy techniques may be best suitable for non-recreational use, whereas the First Person technique may be more suited for entertainment purposes. Of course, it remains to be seen if other techniques that we did not explore in this study can provide better results and be regarded highly among users. However, from our initial exploration on techniques to control UAVs such as the AR Drone, we find that those which provide the best user experience also leverage metaphors that closely associate with the UAV's nature.

REFERENCES

1. Dean, F. S., Garrity, P., and Stapleton, C. B. Mixed Reality: A Tool for Integrating Live, Virtual & Constructive Domains to Support Training Transformation. In *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)*, vol. 2004, NTSA (2004).

2. Fikkert, W., van der, P. V., and Nijholt, A. Gestures in an intelligent user interface. In *Multimedia Interaction and Intelligent User Interfaces: Principles, Methods and Applications*, L. Shao, C. Shan, J. Luo, and M. Etoh, Eds., no. Multim in Advances in Pattern Recognition. Springer Verlag, London, September 2010, 215–242.

3. Gadea, C., Ionescu, B., Ionescu, D., Islam, S., and Solomon, B. Finger-based gesture control of a collaborative online workspace. In *Applied Computational Intelligence and Informatics (SACI), 2012 7th IEEE International Symposium on* (may 2012), 41 –46.

4. Guo, C., and Sharlin, E. Exploring the use of tangible user interfaces for human-robot interaction: a comparative study. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, ACM (New York, NY, USA, 2008), 121–130.

5. Holm, S. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics 6* (1979), 65–70.

6. Lambrecht, J., Kleinsorge, M., and Kruger, J. Markerless gesture-based motion control and programming of industrial robots. In *Emerging Technologies Factory Automation (ETFA), 2011 IEEE 16th Conference on* (sept. 2011), 1 –4.

7. Lichtenstern, M., Frassl, M., Perun, B., and Angermann, M. A prototyping environment for interaction between a human and a robotic multi-agent system. In *Human-Robot Interaction (HRI), 2012 7th ACM/IEEE International Conference on* (march 2012), 185 –186.

8. Ng, W. S., and Sharlin, E. Collocated interaction with flying robots. In *RO-MAN, 2011 IEEE* (31 2011-aug. 3 2011), 143 –149.

9. Nguyen, T. T. M., Pham, N. H., Dong, V. T., Nguyen, V. S., and Tran, T. T. H. A fully automatic hand gesture recognition system for human-robot interaction. In *Proceedings of the Second Symposium on Information and Communication Technology*, SoICT '11, ACM (New York, NY, USA, 2011), 112–119.

10. Sugiura, Y., Kakehi, G., Withana, A., Fernando, C., Sakamoto, D., Inami, M., and Igarashi, T. Walky: An operating method for a bipedal walking robot for entertainment. In *ACM SIGGRAPH ASIA 2009 Art Gallery & Emerging Technologies: Adaptation*, SIGGRAPH ASIA '09, ACM (New York, NY, USA, 2009), 79–79.

11. Urban, M., and Bajcsy, P. Fusion of voice, gesture, and human-computer interface controls for remotely operated robot. In *Information Fusion, 2005 8th International Conference on*, vol. 2 (july 2005), 8 pp.