User-Defined Gesture for Flying Object

Anonoymous

ABSTRACT

Nowadays, ways of interaction between human and computer are more diversified; however, these smart products are mostly immobile. With the development in the field of robotics and automation, intelligent products are no longer stationary. Although the present study of interactions are appropriate for tablet or Smartphone, they are not yet been fully discussed in the field of flying objects. In this paper, we design a flying prototype that combined aircraft and webcam for human computer interaction called Hummieye. This paper presents a complete user-defined gesture set that relies on eliciting gestures from users including control the flying object, take a picture or videoetc. Also, we use the think-aloud method to understand users mental models when they perform the gestures. Our experimental results discover that the gestures of most users will be limited to the prior experience of using a tablet or Smartphone. Nevertheless, we found that users will use different number of hands or fingers to interact with flying object based on the different distance. Moreover, users provide unexpected gestures when they fine tuning the flying object or asking it to perform some instruction. Our results will provide designers create better ways of interaction for flying object.Our results will help designers create better gesture control sets informed by user behavior.

Author Keywords

AR.Drone; UAV; Gestures; User-defined gesture; Kinect

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation: :User Interfaces - *Interaction styles, evaluation/methodology, user-centered design.*

INTRODUCTION

Unmanned aerial vehicles (UAVs), also known as drones, are aircrafts whose flight is controlled remotely by a pilot or controlled autonomously by computers. A drone is a representation of technological advance and is commonly known for military use to reduce casualties. However, up to date, most people do not possess drones, aircrafts, flying objects, and so forth. Thus, the manipulation over or the interaction with a flying object remains unfamiliar to most people. Such unfamiliarity also leads to a paucity of previous works on human-drone interaction. Although some possessing flying objects

Figure 1. A user performing a gesture to move the Parrot AR.Drone forward.

such as low-price, small-sized, radio-controlled helicopters may have experiences on maneuvering its flight over a controller or a smartphone, the flying objects are usually used for games and entertainments instead of useful productivity and utility such as photography. Moreover, there still exists a steep learning curve to become a skillful pilot of a UAV with a controller or a smartphone. The goal of this research is to cull, classify, and analyze intuitive gestures created by participants and to conclude the most favorable gestures for each corresponding function (e.g., pitch, roll, and yaw) of a UAV. We expect that drone manipulation by user-defined gestures should be fun and greatly reduce the learning curve resulting in improvements of human-drone interaction. Moreover, the result of this work provides drone designers, developers, and engineers an insight into such novel human-drone interactions with expected gestures.

There are three main control axes on the drone, including a longitudinal axis (roll), vertical axis (yaw), and lateral axis (pitch), controlling the drone to move left and right, forward and backward, and rotate in clockwise and counterclockwise, respectively. In addition, the change in rotational speed of the propellers in turn pushes the drone to a new altitude. Due to such hardware design of a drone, we asked participants to come up with gestures for these translational and rotational basic controls.

Photographs are one of the most crucial memorabilia to keep track and share moments of our daily lives. Fortunately, some cheap, small-sized UAVs such as Parrot AR Drone equipped with cameras have made flying objects involve in the field of photography. Moreover, drones without being subject to topographical limitations have the advantage to fly to any position unreachable by man. Thus, users are able to take photos in a way that they never could before. Nevertheless, in order for UAVs to play a role in photography, drones should be easy to use, and its control methods must be natural and intuitive.

We have confidence that gestures or body movements are one of the best control methods for UAVs where improvements on human-drone interaction and flying experience can be made. As a result, this study focuses on user-defined gestures, taxonomy of gestures, and the Likert-scale of gestures.

RELATED WORK

Wobbrock et al.[5]conducted a study of gesture taxonomy for surface computing. Wobbrock recruited 20 participants to perform gestures for several tasks. He collected 1080 gestures and manually classified each gesture along four dimensions: form, nature, binding, and flow. Each dimension contains multiple categories. His work provides important insight into the design of gestures for tablets. However, the gestures and its taxonomy concluded in this work were subject to 2D touch-screen devices without any further discussion involving 3D in-air gestures.

Vafaei [4] offered more detailed dimensions of gesture taxonomy. He proposed two main groups of the dimensions: gesture mapping and physical characteristics. Gesture mapping, including nature, form, binding, temporal, and context dimensions, involves how users map gestures to tasks. The group of physical characteristics focuses on the physical attributes of a gesture itself categorized into six dimensions: dimensionality, complexity, body part, handedness, hand shape, and range of motion. Although Wobbrock et al.'s and Vafaei's works put stress on touch screens where no in-air gestures were found, the idea of the classification of gestures was quite inspirational for our study.

Hansen et al.[2] explored four different gaze control modes to control a drone. Gaze interaction provides accessibility for disabled people. Drones may enable those with low mobility to quickly visually inspect surroundings or quickly obtain objects nearby. Nonetheless, the results of the work indicated that the participants mostly favored the mouse over gaze in terms of ease-of-use and reliability. Therefore, future work of similar topic is expected.

Graether and Mueller [1] provided preliminary insights into the design of a flying robot as a jogging companion. The insights includes four themes: embodiment, control, personality, and communication. Graether and Mueller discovered that the communication between the jogger and the flying robot is best through hand gestures.

Pfeil et al. [3] explored five 3D gesture metaphors for interaction with UAVs, including first person, game controller, the throne, proxy manipulation, and seated proxy. Metaphors of proxy manipulation and seated proxy have the highest rating since the gesture sets are more natural, intuitive and are metaphors of grasping the drone in user's hand. Based on the results of this work, participants using gesture control were capable of completing the trial in less time compared to that using smartphone control. Such result is significant and tremendous for the field of human-computer interaction. However, the gesture sets defined by the author instead of users may not be the most intuitive, natural. We believe that user-defined gestures for drones should do even more improvements on the user experience. returned to you to fix.

USER STUDY

Developing Gesture Set

Our user study was mainly divided into two parts, namely, Study 1 and Study 2. For Study 1, 16 participants were recruited and asked to think aloud and perform in-air gestures for each basic control (e.g. moving forward, backward, left, right, and so forth). After the gestures being developed, we then classified the gesture mappings along several dimensions in reference to the related works. For Study 2, another 16 participants were told to pick their favorite gesture and rate every gesture from each category of basic controls where all the gesture sets were created by the participants from Study 1.

Referents and Signs

Based on the hardware design of Parrot AR Drone 2.0, the 10 basic movement controls included forward, backward, left, and right translation, rotation, ascend, descend, takeoff, landing, hovering. There were also 2 basic camera controls including the activation of picture taking and video shooting. Thus, there was a total of 12 basic controls on the Drone.

Apparatus

Parrot AR.Drone 2.0, a quadrotor vehicle, equipped with a camera facing was chosen as the flying object since it is popular, programmable, and affordable. FreeFlight app is the official smartphone app allowing joystick-like control made by Parrot SA. We made use of an open source Processing project that can control the Drone with keyboard presses. The Processing application ran on a Macbook with OSX 10.9.4, 8 GB RAM, a 2.6 GHz Intel i5 processor.

Participants

32 paid participants with age ranging from 19 to 25 volunteered for the study. None of the participants had experience with piloting a flying object. All participants were right-handed. All were college students recruited from various fields including financial management, law school, chemical engineering, medicine school, and so forth. Half of the participants consisted of 6 females and 10 males attended Study 1 for gesture set development. The other half consisted of 6 females and 10 males took part in Study 2 for gesture ratings.

Trial Design

The trial designed by Pfeil et al. [3] was to ask users to stand in front of a Kinect device, use gestures to fly the Drone to the 4 corners of a rectangular test space, and eventually land on a designated spot within the test space. We designed our trials for Study 2 in a similar manner (FIGURE 2). Our test space was approximately 4 m long, and 4 m wide. The participant was placed away from the area for a full view of the test space. At point 1 was where the Drone began to take off. At point 2 and point 3 were two lightweight columns with 1.2 m and 0.6 m in height, both made of empty beverage cans stacking up. The participant was asked to push the column with the Drone at point 2 and 3 in sequence. Point 4 was a 1.8 m-high obstacle made of stacked grocery boxes, and the Drone should fly over the obstacle. Behind the obstacle

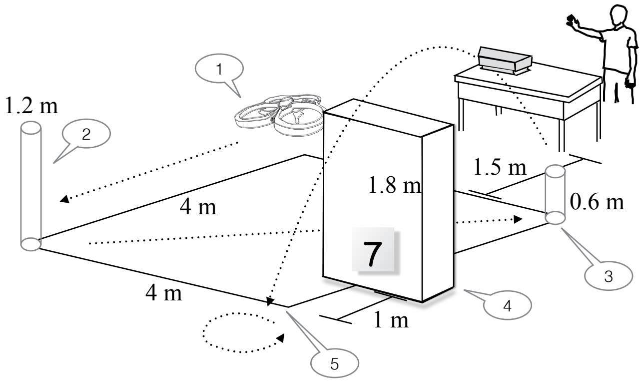

Figure 2. Our test space was approximately 4 m long, and 4 m wide. The participant was placed away from the area for a full view of the test space. At point 1 was where the Drone began to take off. At point 2 and point 3 were two lightweight columns with 1.2 m and 0.6 m in height, both made of empty beverage cans stacking up. The participant was asked to push the column with the Drone at point 2 and 3 in sequence. Point 4 was a 1.8 m-high obstacle made of stacked grocery boxes, and the Drone should fly over the obstacle. Behind the obstacle was a piece of paper written with a number that was not visible from the participant's viewpoint.

was a piece of paper written with a number that was not visible from the participant's viewpoint. The participant should rotate the Drone and read the number through the computer screen receiving live pictures from the camera of the Drone. After reading the number, the Drone landed immediately, and the trial was completed. If the Drone crashed during the trial, the Drone would take off at the crash site and continue the rest of trial. Each participant completed the trial 5 times.

Procedure

In Study 1, an approximate 90-second short video of the introduction and motivation of this project was played in the beginning. Second, another 60-second short video showed how the drone moves. Then, the participant was asked to stand up and start to come up with 12 gestures for all 12 basic controls in random order. In addition, there was one advance movement control, point-to-specified-destination control, we thought this could be exceptionally useful and powerful relative to basic movement control for the Drone to reach a desired destination. Therefore, the participant had total of 13 gestures after Study 1. We allowed the participant to change any previously defined gesture if a new idea came to mind. After all gestures being developed, we employed a Wizard of Oz approach to give the participant an idea of how the gestures they created worked.

After Study 1, we categorized the user-defined gesture sets. Since the gesture for left and right were "symmetric", and so were that for ascend and descend, we grouped them into left-right and ascend-descend gesture sets, respectively, leading to 11 gesture sets. All the actions of all 11 gesture sets were recorded in a video which was played for all participants to pick their favorite gesture for each gesture set in Study 2. Then, the favorite gestures along with Wizard of Oz approach were used in the trial. At last, all participants in Study 2 were asked to watch the same video with all gesture sets and rate for every gesture according to their preference using a 5-point Likert scale.

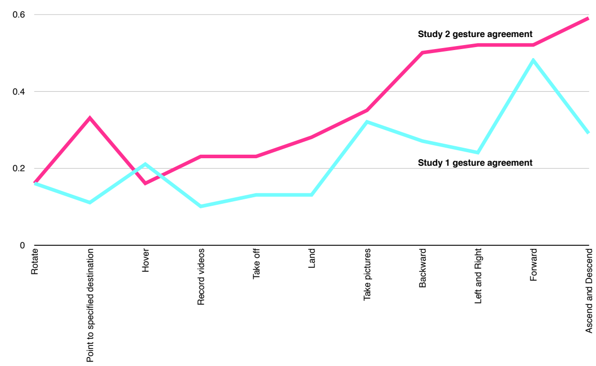

Figure 3. With Caption Below, be sure to have a good resolution image (see item D within the preparation instructions).

RESULT & DISCUSSION

Classification of 3D Gestures

Gesture taxonomy has been used in several works involving speech gestures, sign language, touch screen gestures, and so forth. However, works involving gestures for flying object seemed rare lately. We managed to explore the most intuitive gestures for UAVs and provide future developers insights for better control methods.

In reference to related works, we manually classified each gesture along seven dimensions: form, nature, flow, dimensionality, handedness, hand shape, and trace. All the gestures were performed by participants from Study 1. Each dimension contains multiple categories, shown in Table 2.

The dimension of form describes the motions of all gestures. The categories of this dimension includes static, large-scale motion, and small-scale motion. It is important to find out how many body parts are involved so that future developers will choose appropriate sensor devices to capture the motion.

In the nature dimension, symbolic gestures are visual depictions with no simple equivalent in the world of objects and movements. For example, a time out gesture symbolizes the hovering action. Pantomimic gestures imitate a real meaningful action. For example, a participant would have the arms straight with clenched fist hand shape on both hands and one hand above the other as if turning a steering wheel. A pointing gesture points out the direction where the participant intends the drone to fly. Manipulative gestures have direct relation with real world object interaction. For example, moving the hands forward indicates pushing away the drone. Finally, abstract gestures are with arbitrary mappings. For example, putting a hand on the head to take off the drone would be an abstract gesture.

In the flow dimension, if a gesture is discrete, the gesture must be completed first and subsequently bring out the desired task. For instance, the drone starts to move right after the user's right arm is straightened towards right. A gesture

Referents	Mean	Std.
Ascend and Descend	1.00	0
Forward	1.25	0.50
Left and Right	1.50	1.00
Backward	1.50	0.58
Take pictures	2.25	0.50
Land	2.75	0.96
Take off	3.00	0.82
Record videos	3.00	0.82
Hover	3.25	1.26
Point to specified destination	3.50	1.73
Rotate	3.50	0.58
Mean	**2.41**	**0.79**

Table 1. Summary of user preference between 3 different interaction methods, it provides mean value, standard deviation, 95% confidence interval for mean(Lower Bound and Upper Bound).

would be continuous if the task is performed during gesticulation. The drone rotates around the user as soon as the user faces at a different direction as an example.

In the dimensionality dimension, gestures are classified according to the number of axes involved. The categories include single-axis, double-axis, tri-axis, and six-axis. The descriptions of each category are listed in TABLE X.

The handedness dimension describes the involvement of the user's dominant hand and non-dominant hand for each task. In our study, no participant performed gestures specifically for the non-dominant hand, so the non-dominant category is no needed. Participants performed gestures either with the dominant hand or both hands.

The hand shape dimension describes the configuration, form, and posture of fingers as performing a gesture. Some hand shapes were performed including free hand, ball holding, bent hand, fist, pinch, open hand, index finger, thumb finger, flat hand, ASL-F, ASL-C, ASL-L, ASL-O, ASL-V, and ILY in reference of Vafaei's work [4]. For example, to have the drone moving away from the user is to perform a gesture with a flat hand. Another example would be performing a gesture with a ASL-V hand shape for activation of taking pictures.

The trace dimension describes a variety of shapes traced by the moving parts of a gesture. Traces include straight line, fan shape, cone, hand shape expansion and contraction, and traceless. The traces are shown in FIGURE X.

Agreement

After all 16 participants had provided gestures for each referent for Study 1, we grouped the gestures within each referent such that each group held identical gestures. Group size was then used to compute an agreement score A that reflects, in a single number, the degree of consensus among participants.

$$A = \frac{\sum_{r \epsilon R} \sum_{P_i \subseteq P_r} \left(\frac{|P_i|}{|P_r|} \right)^2}{|R|} \tag{1}$$

Taxonomy of Game Controls		
Form	Static	no motion occurs in G.
	Large-scale	substantial motion occurs in G.
	Small-scale	little motion occurs in G.
Nature	Symbolic	visually depicts a sign
	Pantomimic	imitates real meaningful actions.
	Pointing	points to a specific direction.
	Manipulative	directly manipulates the object.
	Abstract	mapping is arbitrary.
Flow	Discrete	Action is performed after G.
	Continuous	Action is performed during G.
Handedness	Dominant	G. is performed by the user's dominant hand/arm
	Bi-manual	G. is performed using both hands/arms
Hand Shape		free hand, ball holding, bent hand, fist,pinch, open hand index finger, thumb finger, flat hand, ASL-F, ASL-C, ASL-L, ASL-O, ASL-V, ILY
Trace	Straight line	Moving parts trace a straight line.
	Fan shape	Moving parts trace a fan shape.
	Cone	Moving parts trace a cone.
	Expansion/ Contraction	Hand posture changes with expansion or contraction.
	Traceless	G. is static.

Table 2. "G." means "Gesture"

In Eq. 1, r is a referent in the set of all referents R, Pr is the set of proposed gestures for referent r, and Pi is a subset of identical gestures from Pr. The range for A is [—Pr—-1, 1]. As an example, consider agreement for forward in study 1 and study 2. Both had five groups of identical gestures. The former had groups of size 1, 10, 2, 1, and 2; the latter of size 0, 11, 3, 1, and 1. For study 1, we compute

$$A_{forward} = \left(\frac{1}{16}\right)^2 + \left(\frac{10}{16}\right)^2 + \left(\frac{2}{16}\right)^2 + \left(\frac{1}{16}\right)^2 + \left(\frac{2}{16}\right)^2 = 0.48 \tag{2}$$

Mental Model Observations

1. Serveral users make different gestures during user experience from they defined early.

2. Some gestures are often maked and take more time to complete which users feel uncomfortable during user experience. For instance, some users decide to use two hands to gesture ASL-L for task "hover". However, after real user experience beginning, hovering is often gestured. Therefore, user would find that continuously making this gesture is quite laborious. If user has decided to gesture fist to define "hover", it would have been more easy and comfortable. Hence, user- defined gesture before user experience probably isn't the real user's favorite gesture. Conversely, after user experience, users will define gestures which is closer to what users really eager in their heart.

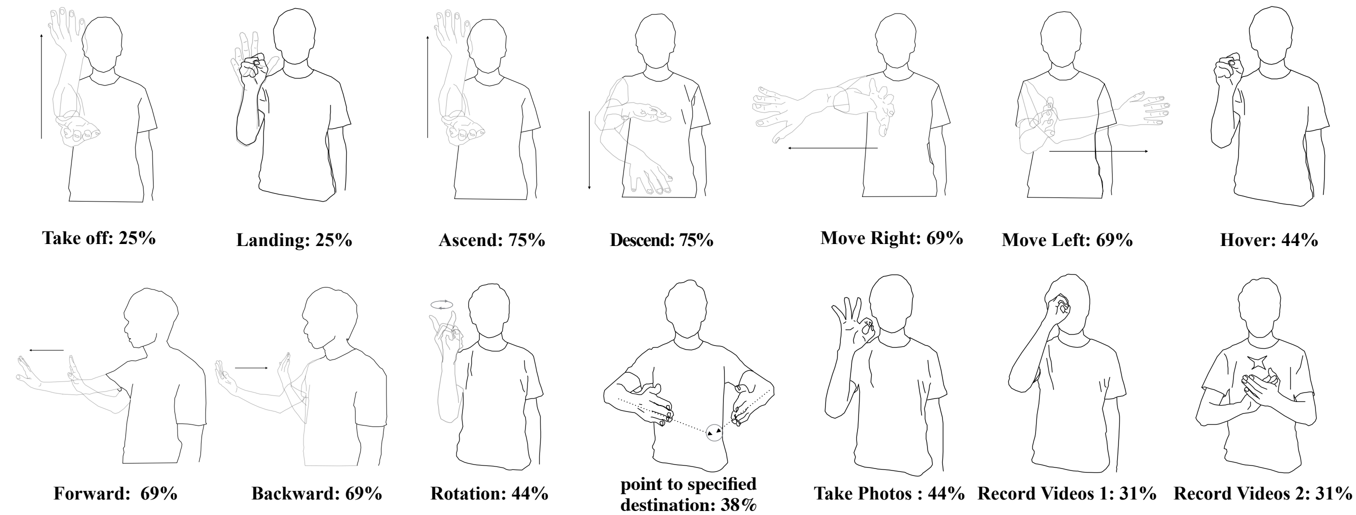

Figure 4. The user-defined game control set with *In-Air* interaction, The percentage indicate the portion of users who perform the control action for the game task.

CONCLUSION

It is important that you write for the SIGCHI audience. Please read previous years' Proceedings to understand the writing style and conventions that successful authors have used. It is particularly important that you state clearly what you have done, not merely what you plan to do, and explain how your work is different from previously published work, i.e., what is the unique contribution that your work makes to the field? Please consider what the reader will learn from your submission, and how they will find your work useful. If you write with these questions in mind, your work is more likely to be successful, both in being accepted into the Conference, and in influencing the work of our field.

ACKNOWLEDGMENTS

We thank CHI, PDC and CSCW volunteers, and all publications support and staff, who wrote and provided helpful comments on previous versions of this document. Some of the references cited in this paper are included for illustrative purposes only. **Don't forget to acknowledge funding sources as well**, so you don't wind up having to correct it later.

FUTURE WORK

The study of this work mainly focused on gesture itself, and several most favorable gestures in each gesture sets were discovered. We plan on implementing a system that can take gestures as input and output basic control commands to a UAV. The input devices may involve Microsoft Kinect for Windows v2 or wearable devices such as a smart watch, a smart band, or smart glasses for motion capture. After the system being built, we also want to compare the completion time of the trial designed in Study 2 between gesture control and smartphone app control. We are also interested in developing features of the camera in the future so that drones can soon play a role in photography in our daily lives.

REFERENCES

1. Graether, E., and Mueller, F. Joggobot: A flying robot as jogging companion. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '12, ACM (New York, NY, USA, 2012), 1063–1066.

2. Hansen, J. P., Alapetite, A., MacKenzie, I. S., and Møllenbach, E. The use of gaze to control drones. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA '14, ACM (New York, NY, USA, 2014), 27–34.

3. Pfeil, K., Koh, S. L., and LaViola, J. Exploring 3d gesture metaphors for interaction with unmanned aerial vehicles. In *Proceedings of the 2013 International Conference on Intelligent User Interfaces*, IUI '13, ACM (New York, NY, USA, 2013), 257–266.

4. Vafaei, F. Taxonomy of gestures in human computer interaction (2013).

5. Wobbrock, J. O., Morris, M. R., and Wilson, A. D. User-defined gestures for surface computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, ACM (New York, NY, USA, 2009), 1083–1092.