Entertainment Multi-rotor Robot that realises Direct and Multimodal Interaction

Kensho Miyoshi
The University of Tokyo
Tokyo, Japan

miyoshikensho@ailab.t.u-tokyo.ac.jp

Ryo Konomura
The University of Tokyo
Tokyo, Japan

konomura@ailab.t.u-tokyo.ac.jp

Koichi Hori
The University of Tokyo
Tokyo, Japan

hori@computer.org

We explore direct interaction of human and multi-rotor robots and its applications for entertainment. In this paper, we present our system that realises direct and multimodal interaction using onboard cameras and a microphone. With these onboard sensors to detect human actions, the robots' reaction chains and expands one after another. In addition, as all the processing is executed within the onboard computer, there is no need to use external devices. We describe its interaction scenario from take-off to landing, and present the pilot evaluation of our system.

Multimodal, Interaction, Multi-rotor robot, Entertainment

1. INTRODUCTION

Multi-rotor robots have recently become one of the most popular materials for hobby and entertainment. Today there are online DIY (Do-It-Yourself) communities¹ where more than fifty thousand people share their knowledge and information on their own aerial robots. While traditional aerial entertainment was “one-way” style that people only control aerial robots with controllers, recent advancement in interactive technologies has enabled “bidirectional” entertainment that people can interact with aerial robots intuitively. However interactive multi-rotor robots often require external sensor devices to detect human actions, which still ends up in situations that human simply make actions toward external sensors rather than directly toward the robots.

In this paper, we present our multi-rotor robot (quadcopter) that realises *direct* and multimodal interaction using onboard cameras and a microphone. As all the processing is executed within the onboard computer, there is no need to use external devices such as controllers, motion tracking system or wireless network system. We introduce the concept of direct and multimodal interaction and introduce our prototype system. Additionally we describe the pilot evaluation.

2. RELATED WORKS

In the past, most of the multi-rotor robots were controlled using controller devices. However, there have recently appeared some approaches to control such robots by human's motion or voice

instead of the devices. Ng made methods to interact with AR.Drone² using hand gestures in front of Microsoft's Kinect [1]. Sanna presented a NUI (natural user interface) framework for quadcopter control [2]. In these cases, stationary cameras are used to detect human motions. Therefore, it is possible to control the robots regardless of the position of the robots. Monajjemi realised an interaction using AR.Drone and its onboard camera to track gestures [3]. As the cameras are attached to the drones, human's motion has to be in the view of the onboard cameras in order to control the drones. Accordingly human motion is induced by the position of the drones. Still, the system is only available within the range of wireless network.

Quigley used PDA (Personal Digital Assistant) as a voice control interface of aerial robots [4]. As mobile microphones are used to detect human's voice, human can control the robot wherever the robot is, or no matter how far away it is.

Figure 1: *Multimodal and Interactive Quadcopter*

¹ DIY Drones, <http://diydrones.com>

² AR.Drone, <http://ardrone.parrot.com>

Figure 2: Left: Interaction using Offboard (Stationary) Camera, Middle: Interaction using Offboard (Mobile) Microphones, Right: Interaction using Onboard Sensors

3. OUR APPROACH

In order to control robots, previous systems used external sensor devices such as motion tracking system, microphones and external network stations. When using external sensor devices, human make actions toward them rather than toward the robots themselves. Stationary network stations limit the available range of robots. Thus, previous multi-rotor robots that at least depend on some external devices have difficulties in realising one-to-one, or *direct* relationship with human. In this chapter, firstly we describe our concept of *direct* interaction of human and robots. Secondly we introduce our prototype system and illustrate its multimodality.

3.1 Direct Interaction

3.1.1. Concept of Direct Interaction

We define *direct* interaction as an interaction between human and a robot that is totally independent of any external devices such as sensor devices to sense human actions or network equipment. The absence of external devices has much potential to remove the limitation of available space of interactive flight, to make aerial entertainment more accessible for people and to explore unique interactivity between human and aerial objects.

Interactive multi-rotor robots are categorized into three groups according to the ways to sense human actions. First is to use offboard and stationary sensor machines such as Microsoft's Kinect (Figure 2, left). Second is to use offboard and mobile machines like Quigley's approach (Figure 2, middle). Third is to use onboard sensors like our approach (Figure 2, right). Figure 2 shows the schematic images of these groups. The yellow areas on the figures below are the areas where human actions can be sensed by sensor devices.

Offboard sensors are useful to actions of particular human, occasionally in particular space. For

example, Microsoft Kinect is capable of accurately sensing actions of human just in front of it. Mobile microphones are useful to sense the voice or sound of the owners. As information obtained by offboard sensors is mostly transmitted to other machines or robots through wireless network, the available range of the systems is limited by the network systems.

It might be useful to use offboard sensors because such interaction is independent of the place and the condition of the multi-rotor robots. Humans can tell their actions to (or, control) the robots no matter how far they are, however loud noise they are making or whichever direction they are facing. As far as the condition between human and sensor devices does not change, the sensitivity of their actions similarly does not change. But this style of interaction ends up in that human is rather "controlling" the aerial robots than "interacting with" them. Each reaction of robot could be separate.

In contrast, as our approach, multi-rotor robots with onboard sensors react to human actions that can be sensed by robots themselves, mostly actions that appear around themselves. The interaction depends much on positional relation between human and robots. The closer the robots get to human, the easier they react to actions of human. Therefore if they react to human actions and move to another place, then they would react to human actions observable from the next place – that is, occasionally to react to another human. There arises a need for humans to change their positions or actions in order to continue interacting with robots. In consequence each reaction of robots induce the actions of humans, which chains one after another because of the direct, or one-to-one relation of humans and robots. In this case, the robots are designed as if being a living creature performing actively against external environment. We believe that this style of interaction would produce a new experience of interacting with aerial robots.

Figure 3: “Passing” the quadcopter (Timeline: from left to right)

3.1.2. Prototype Implementation

Figure 1 shows our prototype of multimodal and interactive quadcopter with two onboard cameras and a microphone based on Konomura’s quadcopter [5]. It is capable of flying and reacting to human actions completely independently of external devices.

Our quadcopter is almost palm-sized and suitable for indoor flight. The combination of MCU (Micro Control Unit) and FPGA (Field-programmable Gate Array) realises its fast image processing, self-localization and stable hovering without use of external control. It measures 12 cm from motor to motor on a diagonal, and the weight is 70 gram. We describe our system in detail from the viewpoints of directness and multimodality.

3.2 Multimodality

Our quadcopter is designed to be multimodal to realise direct and smooth interaction from take-off to landing. Take-off and landing is operated with sound, and the movement in flight is controlled with visual information, especially about the motion of hands wearing gloves.

3.2.1. Vision-Based Approach

Our prototype quadcopter has two onboard cameras. One is attached at the bottom, and the other in front. Using the camera attached at the bottom, the quadcopter is capable of flying and following above people’s hands wearing gloves of a particular colour. The gloves help the quadcopter find the hand areas. We use OpenCV library to detect the colour (Figure 4). The process of extracting colours is executed in about 10 fps.

Figure 4: Hand tracking using OpenCV. Right: Raw Image, Left: Extracted Hand Area

Without hands, the quadcopter keeps hovering at its place by detecting the feature points of its environment [6], which helps a great deal in making

smooth interaction and in avoiding bumping into surroundings.

3.2.2. Audio-Based Approach

We use Julius [7], an open-source voice recognition engine to detect voice commands. Julius is capable of executing real-time voice recognition and open for customizing its library and language model. We made our original library with a set of command words and a name of our quadcopter.

When propellers are rotating, there appears a big noise that prevents from detecting voice commands. Since the noise of propellers mostly distribute in frequency from 100Hz to 1kHz, it is hard to interact with the robot using voice. Meanwhile, there is comparatively small noise from 2kHz to 3kHz. We use a 3kHz whistle to realise sound recognition even when the quadcopter is in the air. With use of FIR (finite impulse response) filter, we have succeeded in extracting the sound of whistle from the total sound that includes the sound of rotating propellers (Figure 5). Owing to this, audio-based communication between human and aerial robots became possible even when they are flying [8].

Figure 5: Sound Spectrum. Above: Sound of Propellers. Below: Sound of Propellers and Whistle

4. INTERACTION SCENARIO

We regard the sequence from take-off to landing as one cycle of interaction, and classify one cycle into three scenes – *take-off*, *flight* and *landing*. In this chapter, we describe how human can interact with our prototype multi-rotor robot in each scene.

4.1 Take-off

Using the voice recognition function, the quadcopter can react to the voice commands and fly up to a certain height to start hovering. Just calling the name can make it take off.

4.2 Flight

When the robot detects the hands with the gloves, it follows and flies above them. It is capable of controlling its height by using the onboard ultrasonic sensor or by processing the images obtained by the front-facing camera. Since the quadcopter is programmed to fly above the maximum area of the selected colour, you can pass the robot to another person just like passing an object hand to hand (Figure 3). Therefore, it is easy for multiple people to join the interaction at the same time because the robot is totally independent of external devices or host machines. Nobody has to hold or wear electric devices.

4.3 Landing

When a whistle is blown, the quadcopter detect it and start landing.

As described above, there is completely no need to use external device such as controllers to make interaction with the quadcopter.

6. PILOT EVALUATION

We organized a pilot evaluation of our system. It was held in an indoor environment. Four persons took part in the experiment and observation – two are not well informed about the system, and the other two are familiar with it. We observed participants' actions.

- (i) The participants without much prior information sometimes moved their hands vertically as if to control the height of the robot. We estimate that the interaction style of the quadcopter in horizontal directions evoked a sense of controlling the height in a similar way.
- (ii) Every participant sometimes failed to induce the robot, and it just kept hovering at its place. It is estimated that they had no means to tell whether the quadcopter detects their hands or not because there was no feedback.
- (iii) From the beginning, every participant always kept looking at the quadcopter in flight. They rarely confirm the positions of their hands. It is estimated there was little need for them to get used to the method to interact with the quadcopter.

Besides these observations, we found some features of our system. Sometimes the quadcopter

oscillated and became unstable in the sky just after it starts to hover still, which happened in areas from 2 to 4 meters away from walls or objects in the room. We expect that this happened because the quadcopter could detect few numbers of feature points in its environment.

7. CONCLUSION

This paper presented our entertainment multi-rotor robot that realises direct and multimodal interaction without any external devices. Since all the processing is executed within onboard computers, it does not require wireless communication either. We also presented the pilot evaluation of our system and gained clues to realise smoother interaction.

This system will not only enhance the availability of interactive flight but also create a new style of aerial entertainment. We plan to explore new styles of interaction between human and aerial robots and applications in entertainment field.

8. REFERENCE

- [1] Ng, W., et al. (2011) Collocated interaction with flying robots. In RO-MAN/IEEE, pp.143-149, 2011
- [2] Sanna, A. et al. (2012) A Kinect-based natural interface for quadrotor control. In Intelligent Technologies for Interactive Entertainment, 2012
- [3] Monajjemi, V., et al. (2013) HRI in the sky: Creating and commanding teams of UAVs with a Vision-mediated Gestural Interfaces. In Intelligent Robots and Systems (IROS), pp.617-623, 2013
- [4] Quigley, M., et al. (2004) Semi-autonomous human-UAV interfaces for fixed-wing mini-UAVs. IEEE/RSJ International Conferences on IEEE, pp.2457-2462, 2004
- [5] Konomura, R., Hori, K. (2013) Designing Hardware and Software Systems Toward Very Compact and Fully Autonomous Quadrotors. In International Conference on Advanced Intelligent Mechatronics (ASME'13), 2013
- [6] Konomura, R., Hori, K. (2014) Visual 3D Self Localization with 8 Gram Circuit Board for Very Compact and Fully Autonomous Unmanned Aerial Vehicles. In 2014 IEEE International Conference on Robotics and Automation (ICRA'14), 2014
- [7] Lee, A., Kawahara, T. (2009) Recent development of open-source speech recognition engine Julius. APSIPA, pp.131-137, 2009.
- [8] Miyoshi, K., Hori, K. (2014) Onboard Sound Recognition using Small Aerial Robot, In proceedings of The Japanese Society for Artificial Intelligence (JSAI'14), 2014 (in Japanese)