키넥트 센서의 음성인식을 이용한 UAV 비행제어

양진영*, 김석훈**, 김기원*

요약

오늘날 키넥트 센서를 이용한 연구가 활발하게 이루어지고 있다. 키넥트 센서는 칼라 영상과 깊이 정보를 제공하므로 활용 분야가 다양하다고 할 수 있다. 키넥트 센서로부터 제공되는 음성인식 기능을 사용하면 게임이나 인간-컴퓨터 상호작용 응용 프로그램 개발을 쉽게 만들 수 있다. 본 논문에서는 마이크로소프트사의 키넥트 센서를 사용하여 인간의 음성을 인식하고, 인식된 결과에 따라 무인항공 로봇을 제어하는 시스템을 구축하고자 한다. 본 논문에서는 12가지 단어를 통해 무인항공 로봇을 제어하는 환경을 구축한다. 최종적으로 각 단어에 대한 인식률을 통해 결과를 분석하였다. 실험에 사용된 12개의 단어를 각각 20회씩 발음하고 각 단어별 인식결과를 분석한 결과 비교적 신뢰할 수 있는 89% 결과를 얻을 수 있었다. 본 실험은 한국인이 영어 인식기능을 사용한 것이며, 만약 한국어 음성인식 기능이 키넥트에서 제공된다면 보다 정확한 인식률을 얻을 수 있을 것으로 보인다.

Flight Control of UAV using Speech Recognition of Kinect Sensor

Jin-Young Yang*, Seok-Hun Kim**, Gi-Weon Kim*

ABSTRACT

Today, research has been actively using the Kinect sensor. Kinect sensor provides color images and depth information that can take advantage of a variety of sectors. Kinect sensor is an essential contributor to game development. It becomes the link between humans and their computers. In this paper explains the system that controls UAV drone robots. Using Microsoft's Kinect sensor, the drone can be controlled by a recognized human voice. This paper presents the 12 words that can control the UAV. It analyzes the result according to the recognition rate of each word. 12 words are pronounced 20 times a piece and each word's recognition result is analyzed with the reliability rate of 89%. In this specific experiment, the user's first language, Korean, was non-applicable as Kinect does not yet recognize the language. Instead, English was used for all voice recognition commands. With a Korean recognition system, a more reliable recognition rate is expected.

Key Words : Kinect sensor, Voice recognition, UAV, Kinect for windows, NUI

* 초당대학교 컴퓨터과학과(✉jyyang@chodang.ac.kr)
** 수원여자대학교 디지털미디어과
· 제1저자(First Author) : 양진영 · 교신저자(Correspondent Author) : 김기원
· 접수일(2012년 11월 12일), 수정일(1차 : 2012년 12월 11일), 게재확정일(2012년 12월 18일)

Ⅰ. Introduction

Recently, an interface for communication between humans and computer has been developed. GUI surrounding in common has changed into NUI surrounding. Among the NUI surroundings voice reception interface based on the sensor, has been focused since it provides a more convenient environment for its users.[1]

This paper explains the system that controls drone robots. Using Microsoft's Kinect sensor, the drone can be controlled by a recognized human voice. Kinect sensor is an essential contributor to game development. It becomes the link between humans and their computers. This paper presents the 12 words that can control the UAV. It analyzes the result according to the recognition rate of each word.

Ⅱ. Theoretical Background

2.1 Kinect

1) Kinect sensor

Microsoft has created a motion sensor add-on for the Xbox 360 gaming console. Kinect, identifies individual players through facial and voice recognition. It includes a 3-D depth camera, which creates a skeletal image of the user and a motion sensor detects their movements. Microsoft's speech recognition software allows the system to understand spoken commands and gesture recognition enables the tracking of player motions.

Kinect's sensor is a horizontal bar connecting a small base to a motorized pivot and is especially designed to be positioned lengthwise above or below the video unit. Attached is an RGB camera that stores three channel data and captures a color image in 1280*960 resolution. An infrared (IR) emitter and an IR depth sensor are also equipped. The emitter sends out infrared light beams and the depth sensor reads the IR beams reflected back from the user.

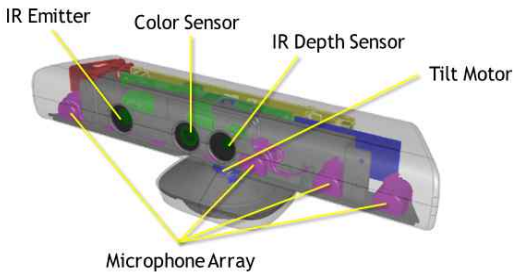

그림 1. 키넥트 센서 구성요소
Fig. 1. Kinect for Windows Sensor Components

The reflected beams are analyzed into depth information, measuring the distance between an object and the sensor. This makes capturing a 3-D image possible. A multi-array microphone, which contains four microphones for capturing sound, is used to record audio as well as find the location of the sound source and the direction of the audio wave.[2]

2) Kinect for Windows

Kinect's software development kit for Windows 7 was released by Microsoft on June, 2011. The SDK includes Windows 7 compatible PC drivers for Kinect device and provides Kinect capabilities for software developers interested in building applications with C++, C#, or Visual Basic by using Microsoft Visual Studio 2010 and includes the

following features.

Raw sensor streams - Access to low level streams from the depth sensor, color camera sensor, and four-element microphone array.

Skeletal tracking - The capability to track the physical image of one or two people moving within the Kinect field of view for gesture driven applications.

Advanced audio capabilities - Audio processing capabilities include sophisticated acoustic noise suppression and echo cancellation, beam formation to identify the current sound source, and integration with the Windows speech recognition API.[3]

To use the rich form of Kinect based natural input, SDK provides a software library and tools to help developers with senses and reactions to real-world events. The Kinect architecture interacts with your application, as shown in <Figure 2>.

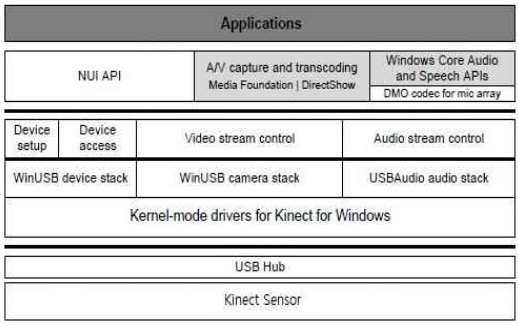

그림 2. 키넥트 SDK 구조
Fig. 2. Kinect SDK Architecture

2.2 UAV

The UAV (Unmanned Aerial Vehicle) is an aircraft with no pilot on board. UAVs can be a remote controlled aircraft or can fly autonomously based on pre-determined flight plans or more even more complex automation systems.

UAVs perform a wide range of functions. The majority of these functions are some form of remote sensing. These include electromagnetic spectrum sensors, gamma ray sensors, biological sensors, and chemical sensors. A UAV's electromagnetic sensors typically include visual spectrum, infrared, or near infrared cameras as well as radar systems. Although uncommon, other electromagnetic wave detectors such as microwave and ultraviolet spectrum sensors may also be used. Biological sensors are capable of detecting the airborne presence of various microorganisms and other biological factors. Chemical sensors use laser spectroscopy to analyze the concentrations of each element in the air.

UAVs usually fall into one of five functional categories. Target and decoy, providing ground and aerial gunnery at a target simulating enemy aircraft or missile. Reconnaissance, providing battlefield intelligence. Combat, providing attack capability for high-risk missions. Research and development, used to further develop UAV technologies to be integrated into field deployed UAV aircraft. Civil and Commercial UAVs, specifically designed and applied for civil and commercial use.[4]

2.3 UAV SDK

Using the SDK provided by the Parrot company, we will build a UAV control system. UAV SDK allows one to easily write your own applications to remotely control the UAV. <Figure 3> is an example of the layered architecture of a host application built upon the UAV SDK.

The current UAV Library provides an open source library with high level APIs to access the UAV. The Library contains set of tools to easily manage the UAV, like an AT command sending loop and thread, a navigation data receiving thread, a ready to use video pipeline, and a ready to use main function.

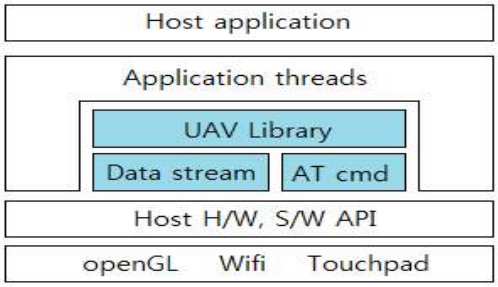

그림 3. UAV 계층적 구조
Figure 3. Layered architecture

All the functions you can call to actually control the UAV are contained in the AT command. Most of them directly refer to an AT command which is then automatically built with the right syntax and sequencing number, and forwarded to the AT management thread.

2.4 Related works

Wenkai et al.[5] explains a study on natural user interface(NUI) in human hand motion recognition using RGB color information and depth information from Microsoft's Kinect camera. To realize this goal, hand tracking and gesture recognition must have no major dependencies of the work environment, for example, lighting or users' skin color. Particular uses in natural interaction with the Kinect device is stored, which then serves to provide RGB images of the environment and the depth map of the present scene. .

Kato et al.[6] proposes an intuitive real time robotically controlled system using human body movement. Recent innovations include motion generation for humanoid robots with reflecting human body movement, which in turn is measured by a motion capture camera. Yet, in existing studies of robot controlled systems by human body movement, the structural information of a robot, for example, degrees of freedom and the range of motion and forms, must be examined to accurately calculate inverse kinematics. Two neural networks compose the related motion generation system: nonlinear principal component analysis and Jordan recurrent neural network.

Matsui et al.[7] focused not only on a robot's joint angle but also a robot's surface movement. Using a neural network, they proposed the motion generation system for imitating the posture of the appearance between the robot and human. However, their system assumed the joint structure of robots to be identical to that of human beings, thus making it difficult to apply the system to various robots.

III. Flight control for UAV

Due to spinning rotors, the schematics of reaction torques on each motor of a quadrotor aircraft. While rotors 1 and 3 spin in one direction, rotors 2 and 4 spin in the opposite direction, establishing control by yielding opposing torques.

Each rotor creates both a thrust and torque about its center of rotation, as well as a drag force against

the vehicle's direction of flight. If all rotors spin at the same angular velocity, with rotors one and three rotating clockwise and rotors two and four counterclockwise, the net aerodynamic torque, and hence the angular acceleration about the yaw axis is exactly zero, meaning that the yaw stabilizing rotor of conventional helicopters is not needed. Yaw is induced by mismatching the balance in aerodynamic torques.

Angular accelerations about the pitch and roll axes can be operated separately without affecting the yaw axis. Each pair of rotating blades rotating controls one axis, either roll or pitch. Increasing the thrust for one rotor while decreasing it for the other will maintain the torque balance needed for yaw stability and induce a net torque about the roll or pitch axes. Therefore, fixed rotor blades can be made to maneuver the quad rotor vehicle in all dimensions. By maintaining a non-zero pitch or roll angle, translational acceleration is achieved .

3.1 Design of speech recognition interface

Kinect voice recognition is currently supported in Australia, Canada, France, Germany, Ireland, Italy, Japan, Mexico, New Zealand, the United Kingdom and the United States. However, speech recognition for the Korean language is not supported. Going forward, we used the voice recognition feature in English.

The microphone array features four microphone capsules and operates with each channel processing 16-bit audio at a sampling rate of 16 kHz.[8,9]. Below, we design a user interface, as the following <table 1>.

표 1. 음성인식 인터페이스
Table 1. speech recognition interface

Type of interface	Pronounce	UAV action
TYpe1	'take off', 'run'	take off
Type2	'land', 'stop'	land
Type3	'turn left', 'left'	turn left
Type4	'turn right', 'right'	turn right
Type5	'go forward', 'go'	go forward
Type6	'go backward', 'back'	go backward

Therefore, UAV will move left or right, when you pronounce the word 'turn left'(type3) or 'turn right'(type4). The moment you say 'go forward(type5)', the UAV will go forward. Want it to go backward? Simply say 'go backward(type6)' and the UAV will go backward.

3.2 Wifi network and UAV connection

The UAV can be controlled from any client device supporting the Wifi ad-hoc mode. Below is a brief description:

1. the UAV creates a Wifi network with an ESSID usually called drone_xxx and self allocates a free, odd IP address.

2. the user connects the client device to this ESSID network.

3. the client device requests an IP address from the drone DHCP server.

4. the UAV DHCP server grants the client with an IP address.

5. the client device can start sending requests the UAV IP address and its services ports.

The Wifi ad-hoc network can also be initiated by the client. If the drone detects an existing network with the SSID it will join the existing Wifi channel.

The drone provides its clients with a navigation data stream. The navigation data is a mean given to a client application to receive periodical information on the drone's status. The navigation data is sent by the drone from and to the UDP port. Information is stored in binary and consists of several blocks of data referred to as options. Each option consists of a header identifying the kind of information contained in it. For example, a 16-bit integer storing the size of the block, and information stored as 32-bit integers.

표 2. UAV 네비게이션 데이터 구조
Table 2. navigation data structure

Header		32bit, int
Drone state		32bit, int
Sequence number		32bit, int
Option 1	id	16bit, int
	size	16bit, int
	data	16bit, int
Option ...		...
Checksum block	id	16bit, int
	size	16bit, int
	data	32bit, int

Host and client, perform initialization using the navigation data. Once the initialization is done, you will be able to control the flight of UAV. <Figure 4> shows the initialization process for the UAV.

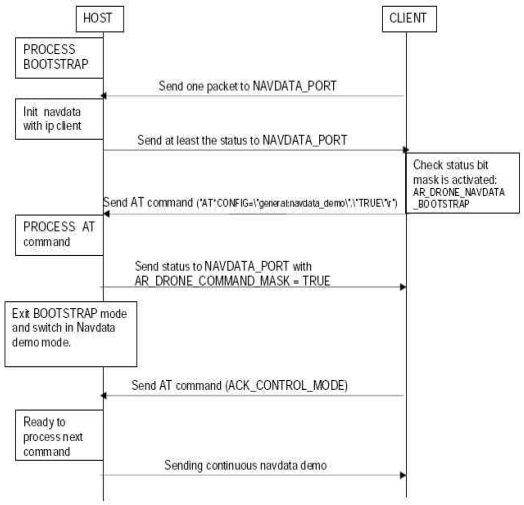

그림 4. 호스트, 클라이언트 초기화
Fig. 4. Host and Client initiation

3.3 Flight Control

The mechanical structure is comprised of four rotors attached to the four ends of an intersection to which the battery and the RF hardware are attached. Each pair of opposite rotors is turning the same way. One pair is turning clockwise and the other anti-clockwise.

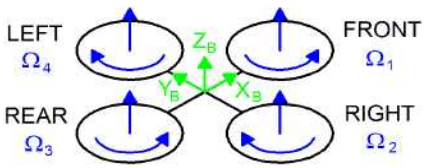

그림 5. UAV 동작 구조
Fig. 5 UAV mechanical structure

Varying left and right rotors speeding the opposite way yields roll movement. This allows the UAV to go back and forth. Therefore, the moment the user

says ‘go forward(type5)’, the UAV will go forward. When the user says ‘go backward(type6)’, the UAV responds accordingly.

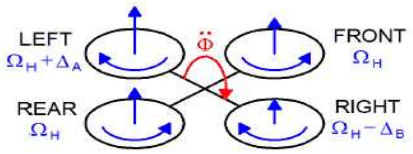

그림 6. UAV 전진, 후진
Fig 6. Roll : UAV go forth and back

Varying front and rear rotors speeding the opposite way yields pitch movement. This allows the UAV to go left and right. Therefore, the UAV will move left or right, when the user pronounces the words ‘turn left’(type3) or ‘turn right’(type4).

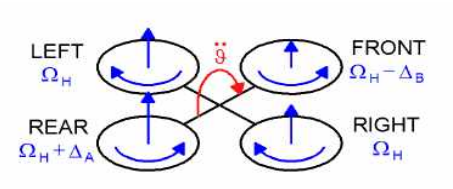

그림 7. UAV 좌측, 우측 비행
Fig. 7 Pitch : UAV go left and right

Ⅳ. Experiments and Results

In this article, we discussed Microsoft’s voice controlled Kinect sensor and its controlling of a drone robot. This experiment’s results are below, in <Table 3>. MS Visual Studio 2010 and Kinect SDK 1.5 were used for developing on an IBM PC platform. The UAV used in this experiment was an AR Drone 2.0 from Parrot company.

표 3. 개발 플랫폼
Table 3. Development platform

Platform	Item
H/W	· Personal Computer · Kinect Sensor · AR Drone 2.0
S/W	· Windows 7 · MS Visual studio 2010 c# · Kinect SDK 1.5 · AR Drone SDK 2.0

<Figure 8>. represents the experiment environment. The Kinect sensor is connected to the computer monitor. The computer monitor displays the vocal recognition program. The UAV for this experiment, named AR Drone II, is seen under the computer monitor.

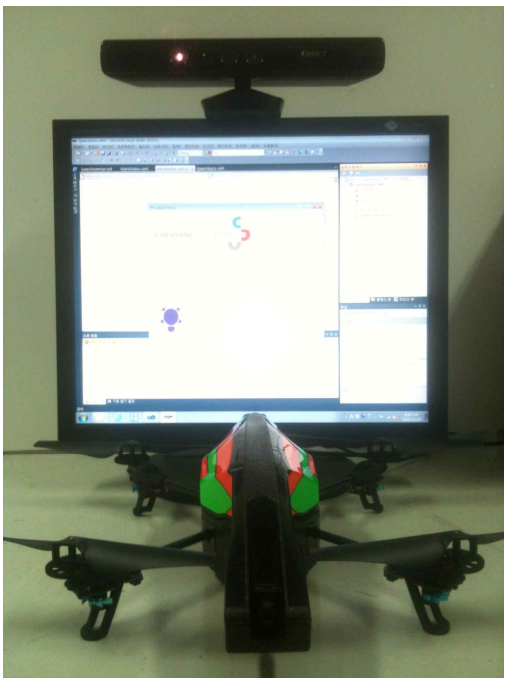

그림 8. 실험 환경
Fig. 8 Experiment environment

The speech recognition experiment results are as follows:

12 words are pronounced 20 times a piece and each word's recognition result is analyzed with the reliability rate.

Table 4 shows the results of speech recognition using the Kinect sensor. In this specific experiment, the user's first language, Korean, was non-applicable as Kinect does not yet recognize the language. Instead, English was used for all voice recognition commands. With a Korean recognition system, a more reliable recognition rate is expected.

In our future work, we plan to continue the research on advanced speech recognition.

표 4. 음성 인식 결과
Table 4. speech recognition result

word pattern	correct	miss	recognition rate
run	20	0	100%
take off	18	2	90%
stop	19	1	95%
land	18	2	90%
left	19	1	95%
turn left	16	4	80%
right	17	3	85%
turn right	16	4	80%
go	20	0	100%
go forward	18	2	90%
back	18	2	90%
go backward	18	2	75%
total	217	23	89.2%

References

[1] I. Oikonomidis, N. Kyriazis, and A.A. Argyros, "Efficient model-based 3D tracking of hand articulations using Kinect," *In British Machine Vision Conference,* pp. 101.1-101.11, 2011.

[2] http://msdn.microsoft.com/en-us/library

[3] Knies, Rob (February 21, 2011). "Academics, Enthusiasts to Get Kinect SDK". *Retrieved* March 18, 2011.

[4] http://www.theuav.com/uav_types.html

[5] Wenkai Xu, Eung-Joo Lee, "Human-Computer Natural User Inter face Based on Hand Motion Detection and Tracking", *Journal of Korea Multimedia Society* Vol. 15, No. 4, April 2012(pp. 501-507)

[6] Akinori Waka bayashi, Satona Motomura, and Shohei Kato, "Associative Motion Generation for Humanoid Robot Reflecting Human Body Movement", *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 12, no. 2, June 2012, pp. 121-130

[7] D. Matsui, T. Minato, K. F. MacDorman, and H. Ishiguro, "Generating Natural Motion in an Android by Mapping Human Motion", *I-Tech Education and Publishing*, 2007.

[8]. V. Pallotta, P. Bruegger, and B. Hirsbrunner, "Kinetic User Interfaces: Physical Embodied Interaction with Mobile Pervasive Computing Systems, in: Advances in Ubiquitous Computing:Future Paradigms and Directions", *IGI Publishing,* February, 2008.

[9] "Kinect for Xbox 360". Microsoft. Retrieved July 7, 2010. "Array of 4 microphones supporting single speaker voice recognition".

저자소개

양진영(Jin-Young Yang)

1983년 조선대학교 경영학과(경영학사)
1988년 조선대학교 전자계산학과(공학석사)
2002년 목포대학교 컴퓨터공학과(공학박사)

1997년 ~ 현재 초당대학교 컴퓨터과학과 교수

※관심분야 : TCP/IP, Traffic Control, MMI

김석훈(Seok-Hun Kim)

2003년 한남대학교 컴퓨터공학과(공학석사)
2006년 한남대학교 컴퓨터공학과(공학박사)

2012년 ~ 현재 수원여자대학교 디지털미디어과 조교수
※ 관심분야 : 모바일컴퓨팅, VoIP, 웹데이터베이스

김기원(Gi-Weon Kim)

1987년 한남대학교 전자계산학과(공학사)
1989년 숭실대학교 전자계산학과(공학석사)
2001년 한남대학교 컴퓨터공학과(공학박사)

1996년 9월 ~ 현재 초당대학교 컴퓨터과학과 부교수
※관심분야 : 비디오 브라우징, 영상처리, 음성인식