

키넥트를 이용한 동작과 음성을 인식하기 위한 컴퓨터 인터페이스 구현

황선명*, 엄희균**, 김범식, 박성주, 임홍택, 이은경, 강진원, 김정섭

*대전대학교 컴퓨터공학과 교수

**대전대학교 컴퓨터공학과 강의전담

대전대학교 컴퓨터공학과

e-mail:dlagmdxor@naver.com

Computer interface construction for recognizing both motion and voice using kinect

Sun-Myung Hwang*, Hee-Gyun Yeom**, Beom-Sik Kim, Seong-Joo Park, Heung-Taek Lim, Eun-Kyeong Lee, Jin-Won Kang, Jeoung-Seop Kim

*Professor, Dept of Computer Engineering, Dae-jeon University

**Dept of Computer Engineering, Dae-jeon University

Dept of Computer Engineering, Dae-jeon University

요 약

현재 컴퓨터 조작하는 방식은 입력장치인 키보드로 조작하는 CUI, 마우스로 조작하는 GUI가 있지만 키보드와 마우스를 사용하기 불편한 장소나 불편한 사람이 동작과 음성만으로 컴퓨터를 조작하는 NUI를 구축하여 보다 나은 컴퓨팅환경을 제공하고 새로운 형식의 콘텐츠를 제공할 수 있는 기틀 마련한다.

1. 서론

최근 들어 스마트폰의 보급과 함께 터치스크린이 가장 편리한 사용자 인터페이스로 각광 받고 있다. 하지만 터치스크린의 한계점은 분명히 존재한다. 그 한계점 중에 한 가지는 여전히 무엇인가를 눌러서 작동하는 방식이다. 터치스크린도 스크린이 키보드 자판이 되고 마우스 버튼이 되는 것이다. 스크린은 눌러진 않지만 누르는 버튼인 것이다. 그 한계점을 고찰해보았다.

부쩍 관심이 높아진 음성인식 기술과 사람의 동작을 이해하는 모션인식 기술을 이용하여 터치 없는 새로운 컴퓨팅환경을 구축하여 누르지 않고 가까이 가지 않아도 작동하는 컴퓨팅환경을 구축한다. 그리고 그 환경에 맞는 콘텐츠를 제공할 수 있는 기틀을 제공하여 기존 사용자 인터페이스에서는 접하지 못한 다양한 콘텐츠를 창조하는 하나의 틀을 마련한다.

2. 보다 나은 성능을 위한 요구사항

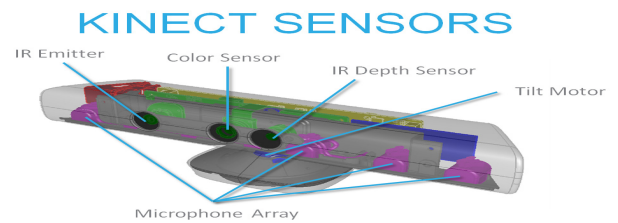
단순히 손의 모양, 사람이 서있는 모습을 캡처하는 것이 아니라 머리, 손, 다리, 등 각 부분별 위치에 따라 다른 동작으로 인식하기 위해 키넥트를 이용하여 단순히 2D카메라로는 입력 받을 수 없는 깊이 정보를 얻어온다.



<그림 1> 2D-RGB카메라 영상(왼쪽)과 깊이 값을 캡처한 영상(오른쪽)

3. 키넥트 구성 및 동작원리

키넥트는 <그림 2>와 같이 RGB카메라, 적외선 인식 센서, 적외선 레이저, 마이크어레이 그리고 틸트 모터로 구성되어있다. 여기서 마이크어레이는 마이크를 배열로 구성해놓은 것이고 틸트 모터는 키넥트의 각도를 조절해주는 모터이다.



<그림 2> 키넥트 구성

사람의 동작을 인식하는 부분은 RGB카메라와 적외선 인식 센서, 적외선 레이저이다. RGB는 단순히 문체의 색이나 크기, 개수를 얻어오며 2차원 데이터를 처리한다. 적외선 인식 센서와 적외선 레이저는 <그림 4>와 같이 적외선 레이저를 이용해 물체의 반사를 적외선 인식 센서가 읽어와 물체의 깊이 정보를 3D로 처리한다. 이 3D 깊이 처리가 키넥트의 핵심적인 기능이다.



<그림 3> 적외선 레이저 작동 모습

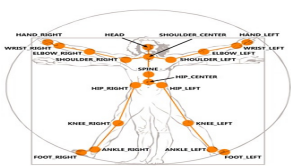


<그림 4> 깊이 정보 프로세싱

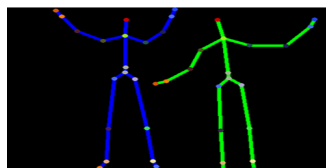
4. 시스템 구현

4.1 모션인식 원리

키넥트로 물체를 인식하고 사람이라면 <그림 5>와 같이 관절을 인식한다. 얻은 관절을 추적하여 <그림 6>과 같이 사람의 뼈대를 그릴 수 있고 관절이나 특정 위치에 가상의 이미지를 띄우거나 따라다니는 물체를 지정할 수 있다. 각 관절마다 x, y, z의 좌표 값이 지정되고 관절의 x, y, z의 좌표 값을 얻어올 수 있다. 이 특성을 이용하여 오른손관절, 왼손관절의 좌표와 화면상 커서의 위치를 매핑을 한다면 우리가 쉽게 사용하는 마우스를 대체하여 사람의 신체로 컴퓨터 제어가 가능해진다.



<그림 5> 인체 각 부위별 관절 부분



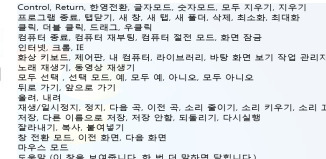
<그림 6> 실제 키넥트로 추적한 사람의 관절을 이용한 뼈대그림

4.2 단축모션

각 관절마다 좌표 값들을 비교하여 관절의 상대적인 값들을 얻어온다. 그리고 그 상대적인 값들을 조건으로 넣으면 특정 모션에 특정 기능을 넣을 수 있다. 오른손관절이 어깨관절보다 위에 있다고 가정하면 지금 사용자는 오른손을 위로 들고 있다고 인식을 시킬 수 있다. 그러나 오른쪽 팔꿈치 관절이 어깨보다 위라고 할 때는 팔을 번쩍 들고 있는 상태임을 알 수 있다. 이처럼 각 관절의 상대적인 좌표를 이용하여 여러 가지 동작을 세밀하게 인식할 수 있다. 그리고 인식한 동작들을 이용하여 다양한 기능을 구현할 수 있다.



<그림 7> 기본 제공 단축 모션 설정



<그림 8> 음성명령어 안내 도움말

4.3 음성인식

음성인식은 마이크로 입력받은 음성소스를 분석하여 미

리 설정해 놓은 음성소스와 매칭이 된다면 알아듣고 수행하는 원리로 진행된다. 그렇기 때문에 음성인식과 관련되어 인식을 향상을 위해 많은 노력이 필요하다. 하지만 아직까지의 음성인식은 위치를 입력받은 음성소스가 우리가 가진 데이터랑 어느 정도 일치하는지 여부로 ‘인식한다.’, ‘인식 못한다.’로 나누어진다. 현재로서 음성인식은 앞으로 데이터가 많아지거나 음성소스에 잡음이나 하울링을 제거하는 방법을 사용하면 보다 나은 음성인식결과를 얻을 수 있다. 물론 새로운 처리 방식이나 새로운 알고리즘을 찾는다면 인식률이 더 좋아질 것이지만 말이다.

영어 인식과 한국어 인식도 비슷한 원리로 작동하는데 키넥트는 한국어 음성을 정식적으로 지원하지 않는다. 그렇기 때문에 한국어 인식은 마이크로소프트에서 제공하는 Speech Platform 한국어를 이용해 키넥트와 별개로 작동할 수 있게 구성해 놓았다.

<표 1> 키넥트를 이용한 응용부분

키넥트 기능	응용부분
관절 추적	손 관절을 추적하여 커서의 움직임을 제어 관절의 좌우 움직임을 계산하여 swipe 모션으로 응용
	관절의 상대적인 좌표 값을 계산하여 단축모션을 구현
음성캡처	음성을 입력받아 다양한 소프트웨어(음성소스 분석 및 후처리 등)와 연동 가능하게 구현
깊이 정보	모션을 2차원이 아닌 3차원으로 적용하여 다양한 모션으로 사용가능하게 구현(단축모션 포함)

5. 결론과 향후 발전방향

키보드와 마우스는 차지하는 공간도 있으며 평평한 곳에 설치를 해야 한다는 점이 있다. 그리고 대형 스크린처럼 근거리에서 전체적으로 볼 수 없는 경우는 불편할 것이다. 그러나 비교적 원거리에서 모션과 음성으로 조작을 한다면 많은 부분 기존의 인터페이스보다 개선되어 사용자에게 다양하고 편리한 사용자인터페이스를 제공할 수 있다. 또한 마이크로소프트 CEO 스티브 발머가 NUI를 차세대 대혁신의 하나로 지목했다. 발머는 지난 2010년 10월 15일 워싱턴대학 컴퓨터공학과 학생들과의 대화에서 “한때는 마우스로 PC를 조작하는 것이 자연스럽게 여겼지만, 지금은 그렇지 않다”라고 했으며, 직접 말을 하거나 키넥트와 같이 동작으로 컴퓨터와 상호작용을 하는 것이 더 자연스러운 방법이라고 강조했다.[1] 이처럼 키보드나 마우스가 사용되어지기 어려운 환경뿐만 아니라 일상생활에서도 키보드와 마우스를 충분히 대체할 수도 있을 것이다.

참고문헌

[1] <http://www.itworld.co.kr/news/63035>에서 일부 발췌
 [2] 찰스 페졸드, “C#과 XAML을 이용한 .NET 3.0 윈도우 프로그래밍 찰스 페졸드의 WPF“ 2007.09
 [3] 박상현, “뇌를 자극하는 C# 4.0 프로그래밍“ 2011.08
 [4] Kinect <http://www.microsoft.com/en-us/kinectforwindows>
 [5] MSDN <http://msdn.microsoft.com/en-us/default.aspx>
 [6] <http://talkingaboutme.tistory.com/category/About%20Kinect>