

키넥트를 이용한 VOD 사용자 인터페이스 설계

임재걸^{*○}, 이강재^{**}, 김행선^{*}

^{*}동국대학교 경주캠퍼스 컴퓨터공학부

^{**}수원과학대학교 컴퓨터정보과

e-mail: {yim, hsc818}@dongguk.ac.kr^{*}, kjlee@ssc.ac.kr^{**}

Design of a User Interface for VOD with KINECT

Jaegeol Yim^{*○}, Kangjai Lee^{**}, Haengseon Kim^{*}

^{*}Dept. of Computer Engineering, Dongguk University

^{**}Dept. of Computer Information, Suwon Science College

● 요약 ●

인터넷 통신량의 대부분이 주문형 비디오(VOD) 서비스임은 이미 널리 알려진 바와 같다. VOD 사용자 인터페이스가 사용자의 손동작을 인식한다면 노인, 유아, 장애인 등 리모트컨트롤의 사용이 쉽지 않은 사용자들도 편리하게 VOD를 즐길 것이다. 게임 등에서는 이미 키넥트(KINECT)를 이용한 사용자 인터페이스가 널리 사용되고 있어, 본 논문은 키넥트를 이용한 VOD 사용자 인터페이스를 제안한다. 제안한 사용자 인터페이스는 키넥트에서 얻을 수 있는 손의 좌표와 깊이 정보 데이터를 이용하여, 사용자가 메뉴를 선택하는 것을 인식하고 사용자가 비디오를 선택하면 해당 동영상 콘텐츠를 재생하는 등의 적절한 동작을 한다. 동영상 콘텐츠는 네트워크 상에서 멀티미디어 데이터 스트리밍 방법에 대한 표준안인 RTSP(Real-time Streaming Protocol)로 전송하여 실시간으로 동영상을 재생한다.

키워드: 키넥트(KINECT), 내추럴 UI, RTSP, 동작인식(Action Detection)

I. 서론

인터페이스 기술의 발전으로 인하여 과거 키보드의 캐릭터 인풋을 기반으로 하는 CLI(Command Line Interface)에서, 주로 마우스를 인터페이스로 하는 GUI(Graphic User Interface)의 시대를 넘어 터치로 대표되는 직관적인 인풋을 사용하는 내추럴 UI(Natural User Interface)의 시대가 됨에 따라 점차 인터페이스의 물리적 접점이 사라지고 있다[1]. 대부분의 컴퓨터 기기들은 키보드나 마우스, 리모트컨트롤 혹은 화면터치로 조작하고 있지만 내추럴 UI는 이러한 조작 환경의 근본을 바꿀 혁신적인 것이다[2].

키넥트는 적외선 카메라와 RGB 카메라는 탑재한 모션 인식 디바이스이다[3]. 키넥트에는 3개의 카메라가 탑재되어 있다. 각 카메라는 적외선을 쏘고, 읽고, 색상을 구분하도록 설계되어있다. 원래 키넥트는 게임용으로 XBOX360의 보조조작 도구로 개발되었지만 키넥트의 가능성을 본 개발자들의 노력으로 마이크로소프트(MS)에서는 윈도우용 키넥트를 발표하게 된다. 이후 동작인식과 관련된 연구가 더욱 활발해지게 되었다.

본 논문에서도 실생활에서 이용 가능한 동작인식 프로그램을 소개한다. 이 프로그램은 스마트 TV 등에서 사용 가능한 어플리

케이션으로 키넥트의 동작인식 기술을 사용하여 사용자의 손동작으로 동영상 콘텐츠의 장르를 구분하고, 그 동영상을 재생하는 기능을 갖추고 있다.

최근 스마트 TV의 보급이 활성화 되면서 기존의 TV에 비해서 양방향 서비스, 주문형 비디오(VOD : Video On Demand), 검색 등 다양한 서비스가 추가돼 이를 조작하기 위해서는 복잡한 UI와 복잡한 조작이 필수적인데, 복잡한 리모컨 조작이 힘든 노년층, 유아, 장애인에게 이러한 내추럴 UI가 많은 도움이 될 것이다.

II. 키넥트 동작 원리 연구

사람의 움직임과 동작을 인식하기 위해서는 깊이 감지 카메라와 RGB 카메라를 이용한 트래킹이 필요하다. 최근 몇 년간 깊이 감지 기술의 발전과 함께, 실시간 깊이 카메라를 이용한 연구들이 진행되고 있다[4, 5]. 특히, 최근에 출시된 마이크로소프트사의 키넥트 센서는 저가의 깊이 카메라로써, 실시간으로 깊이 정보뿐만 아니라 RGB 영상과 관절 추적 정보를 제공한다[6]. 키넥트 센서로 부터 제공되는 데이터(깊이, RGB, 관절 위치)의 사용은 제스처 인식을 위해 필요한 사람/신체부위 검출 및 포즈 추정의 수고를

달어주고, 게임이나 인간-컴퓨터 상호작용 응용 개발을 쉽게 만들고 있다[7, 8].

키넥트는 그림1과 같이 3개의 렌즈로 구성되어 있다. 두 개의 (A)렌즈가 3D 깊이 센서이다. (A)카메라 중 바깥쪽의 적외선 프로젝터에서 적외선을 픽셀 단위로 쏘아주고 다른 적외선 카메라가 반사되어 돌아오는 적외선을 인식한다. (B)렌즈는 RGB 카메라로 RGB 영상을 인식한다.

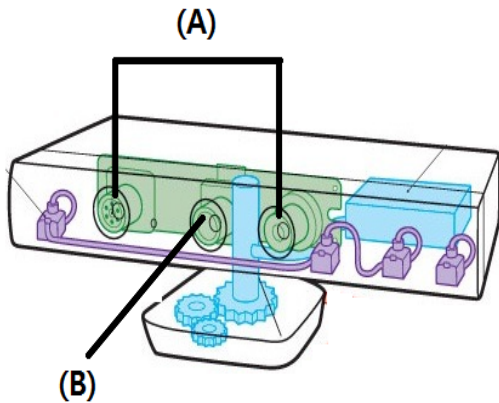


그림 1. 키넥트 구성
Fig. 1. Composition of KINECT

이런 키넥트를 컴퓨터에서 사용하기 위해서는 마이크로소프트사에서 배포하는 키넥트용 SDK(Software Development Kit)가 필요하다. 키넥트 SDK에서 제공하는 Skeleton Tracking을 이용하면 손, 어깨, 팔꿈치, 다리, 무릎 등 사람 신체의 각 부분을 인식할 수 있다. 본 논문에서는 사람의 손을 인식하여 VOD를 조작하는 방법을 제안한다.

III. 내추럴 UI와 멀티미디어 재생

내추럴 UI는 사용자가 직접 손을 뻗는 등의 동작을 하여 편리하게 인터페이스를 조작할 수 있다. 최근 IPTV 사업에 있어서 리모트컨트롤은 IPTV 보급과 성과에 있어서 직접적으로 연관이 있다. 기존 TV에 비하여 양방향 서비스, VOD, 검색 등 다양한 서비스가 추가돼 이를 조작하기 위해 복잡한 UI에 대한 조작이 필수적이다. 하지만 기존 리모트컨트롤은 사선으로의 화살표 이동, 단축 이동 등이 난해하여 UI 디자인 또한 제한이 된다. 따라서 3차원 공간에서 사용자 움직임을 인식해 마우스처럼 자유롭게 포인터 이동, 입력, 제어를 하는 내추럴 UI는 사용자의 조작환경을 혁신적으로 바꿀 수 있다. 또한 리모트컨트롤 조작이 힘든 유아, 노년층, 장애인도 편리하게 TV를 조작할 수 있을 것이다.

본 논문에서 소개할 VOD 사용자 인터페이스는 IPTV 등에서 VOD 서비스를 이용할 때 리모트컨트롤 없이 VOD의 장르를 선택하고 비디오 선택, 비디오 재생까지 사용자의 손동작만으로 위

서비스를 이용할 수 있도록 설계하였다. 키넥트가 사용자의 손을 트래킹하기 시작하면 화면에 손모양의 커서가 표시된다. 사용자가 손을 움직이면 커서가 이동하게 된다. 손을 움켜쥐면 스크롤바를 좌, 우로 이동시켜 원하는 콘텐츠를 찾을 수 있다. 또한 이용하고자 하는 콘텐츠에 커서를 이동하고, 사용자가 손을 앞으로 뻗는 동작을 수행하면 해당 콘텐츠가 선택된다.

선택한 동영상 콘텐츠는 네트워크 상에서 멀티미디어 데이터 스트리밍 방법에 대한 표준안인 RTSP(Real-time Streaming Protocol)로 전송하여 실시간으로 동영상을 재생한다. RTSP는 응용 프로토콜로서 클라이언트 서버환경에서 실시간 특성을 갖는 데이터 전달을 제어하는데 이용된다. 즉, VOD 서버, 오디오 서버, 멀티미디어 서버들로부터 저장된 정보를 실행시키고 이를 제어하기 위한 프로토콜이다. RTSP는 웹 환경에서 사용할 수 있다는 장점을 갖고 있다. 전송은 UDP(User Datagram Protocol), 멀티캐스트 UDP, TCP(Transmission Control Protocol)를 사용하며 RTP를 이용하여 멀티미디어 데이터를 전달할 수도 있다. RTSP는 On-demand 전송과 실시간 데이터(오디오, 비디오)의 제어를 가능하게 하기 위해 HTTP(Hyper Text Transfer Protocol)의 틀을 확장한 것이다. 데이터의 소스는 라이브 데이터 피드와 저장된 클립을 모두 포함할 수 있다. RTSP는 시간적으로 동기화된 연속적인 미디어(오디오나 비디오)를 제어하기 위해 만들었으며, 그 자신이 연속된 스트림을 전달하는 것은 아니다. 단지 스트림 제어가 가능하도록 한 것으로 멀티미디어 서버에 대한 네트워크 원격 제어로 활용하기 위한 것이다[9].

IV. 설계

내추럴 UI에서 사용자의 모션 인식을 가능케 하려면 키넥트와 마이크로소프트에서 배포하는 키넥트용 SDK가 필요하다. SDK에서 제공하는 Skeleton Tracking을 이용하여 사람의 각 신체 중 사람의 손을 인식하여 화면에 손 모양의 커서를 생성하고 사용자의 손 움직임에 따라 커서가 이동하도록 한다. 먼저 키넥트 센서 데이터에서 인식 된 신체 각 부분의 좌표 중 손의 좌표만 따로 추출하여 그 좌표를 트래킹하여 커서를 갱신하면 된다. 이 커서는 키넥트 SDK에서 제공한다. 우리는 커서가 각 콘텐츠 위에서 주먹을 움켜쥐거나 손을 뻗었을 때 생기는 이벤트를 설계하면 된다.

VOD 사용자 인터페이스의 기본적인 흐름은 그림2와 같다. 프로그램이 시작되면 키넥트로부터 사람의 손 좌표를 얻어와 화면에 커서를 표시한다. 장르를 구분할 수 있도록 만든 버튼 위로 커서를 올리고 손을 앞으로 뻗는 동작을 하면 데이터베이스 서버에 담겨져 있는 비디오 리스트에서 선택한 장르의 비디오만을 화면에 출력한다. 각 비디오에는 Video_path가 지정되어 있다. 이 Video_path는 스트리밍을 하기 위한 RTSP 주소이다. 비디오 중 하나를 손을 뻗는 동작을 취해서 선택하면, 해당 비디오가 스트리밍이 된다.

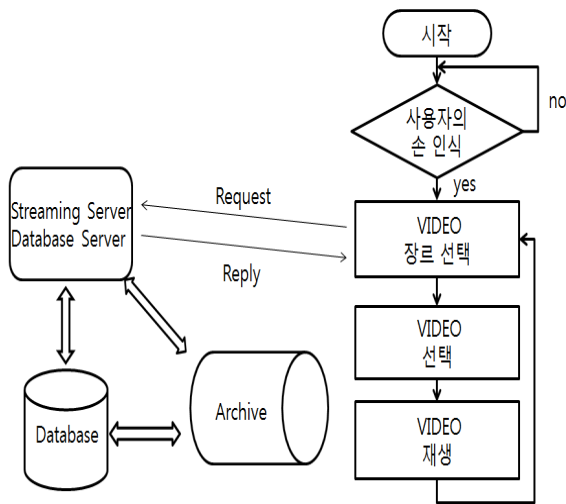


그림 2. 프로세스 흐름도
Fig. 2. The process of the application

단지 VOD 재생만이 아니라 다양한 VOD 서비스를 위한 연구도 필요하며, 현재 진행 중이다.

그림 3은 구축 중인 VOD 서비스를 위한 데이터베이스의 구조를 보인 것이다. 사용자의 인적 사항과 비디오 관람 기록을 저장하여 효율적인 VOD 추천 기능을 추가하고자 한다.

또한 다음과 같은 다양한 웹 서비스들을 개발하여 더욱 편리한 VOD 서비스를 제공하고자 한다.

- 데이터베이스 이용을 돕는 다양한 웹 서비스
- 비디오들을 주어진 기간 동안 보여진 횟수로 내림차순 정렬하는 웹 서비스
- 사용자들을 주어진 사용자와 가장 비슷한 순으로 내림차순 정렬하는 웹 서비스
- 다양한 비디오 추천을 위한 웹 서비스 등

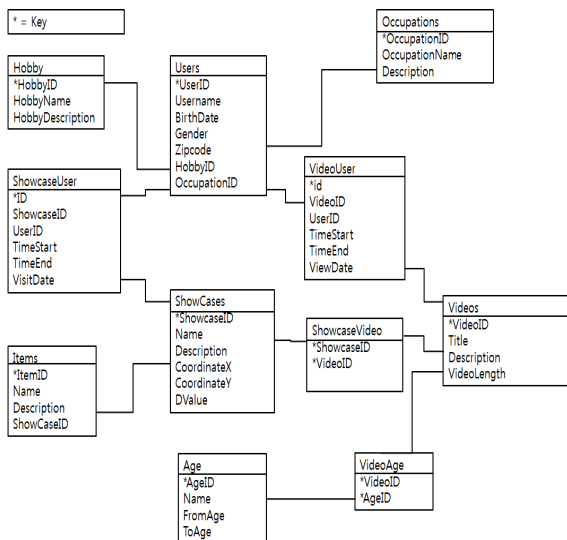


그림 3. VOD 서비스를 위한 데이터베이스 구조
Fig. 3. The database structure for VOD services

V. 실험

본 논문에서 구현한 VOD 사용자 인터페이스를 보면, 그림4의 (A)장르 선택화면과 같이 각 장르별로 구분되어 있는 버튼이 생성된다. 손을 이동시키고 손을 앞으로 뻗는 동작을 취하여 버튼 중 하나를 선택하게 되고, (B)의 동영상 선택화면이 출력된다. 각 비디오는 장르 선택화면에서 선택한 장르의 비디오만 리스트에 생성된다. 사용자는 비디오 중 하나를 선택하는 동작을 취하면 (C)와 같이 비디오가 재생되게 된다. 우측 하단의 Back 버튼을 선택하게 되면 (A)장르 선택 화면으로 되돌아갈 수 있다.

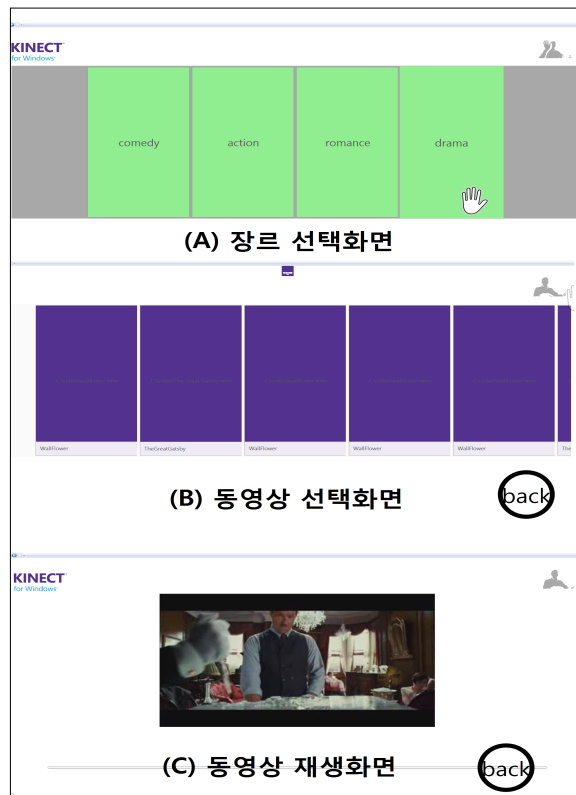


그림 4. 프로그램 실행 화면
Fig. 4. Screenshots

VI. 결론

본 논문은 키넥트의 깊이 카메라 기능으로 다양한 서비스를 갖춘 IPTV의 복잡한 리모트콘트롤을 대체하여 사용할 수 있는 인터랙티브 VOD 사용자 인터페이스를 설계하고, 사용자의 편의성을 향상시켜보자는 목표로 연구하였다. 키넥트와 같이 모션 인식 카메라를 사용하면 사용자가 직관적이고 심플한 UI에서 편리하게 VOD 서비스를 이용할 수 있다.

앞으로 본 논문의 더욱 완성적인 결과를 위해서 아래와 같은 연구를 더 하여야 한다.

첫째로, 키넥트는 음성 인식 기능도 갖추고 있다. 모션 인식과 음성 인식의 통합을 통해 더 편리하게 VOD 서비스를 사용할 수

있는 방법을 연구해야 한다.

둘째로, 손을 뻗는 동작 등 움직임이 큰 동작은 유아나 장애인, 노년층의 사람들은 쉽게 동작을 하지 못하는 경우도 있다. 조금 더 작고, 세밀한 동작으로 조작을 할 수 있는 방법을 연구해야 한다.

이처럼 키넥트에 다양한 영역에서 활용하고자 하는 시도와 연구가 계속되고 있다. 키넥트가 게임뿐만 아니라 사용자의 편의성을 위한 프로그램으로 개발되고, 위와 같은 키넥트를 이용한 다양한 서비스 개발이 이루어지기를 기대한다.

사사

This work was supported by the MKE (Ministry of Knowledge Economy) [A004700008], Development of realistic sense transmission system with media gateway supporting multi-media and multi-device and C0033172 by Business for Cooperative R&D between Industry, Academy, and Research Institute funded Korea Small and Medium Business Administration in 2012.

참고문헌

- [1] S. Oh, Y.-Y. Ham, J. Eune, "Designing Gesture Interaction without Midas Touch Issue : Emphasis on Coverflow UI", HCI 2012, pp.576-578, 2012.1.
- [2] Natural UI, http://navercast.naver.com/contents.nhn?rid=122&contents_id=28416
- [3] H.-J. Jung, H.-K. Lee, "Implementation of Real-Time Streaming Service Using Live555 RTSP Server", KIISE 2012 Fall, pp.188-190, 2012.11.
- [4] V. Ganapathi, C. Plageman, D. Koller, and S. Thrun, "Real Time Motion Capture using a Single Time-of-flight Camera", Int. Conf. on Computer Vision and Pattern Recognition, pp.755-762, 2010.
- [5] M. Siddiqui and G. Medioni, "Human Pose Estimation from a Single View Point, Real-time Range Sensor", Int. Workshop on Computer Vision for Computer Games at Conf. on Computer Vision and Pattern Recognition, pp.1-8, 2010.
- [6] S. Cho, H. Byun, H.-K. Lee, J. Cha, "Hand Gesture Recognition from Kinect Sensor Data", Journal of Broadcast Engineering, Vol.17, Issue 3, pp.447-458, 2012.
- [7] J. Sung, C. Ponce, B. Selman, and A. Saxena, "Human Activity Detection from RGBD Images", Int. AAAI 2011 Workshop, pp.47-55, 2011.
- [8] I. Oikonomidis, N. Kyriazis, and A.A. Argyros, "Efficient Model-based 3D Tracking of Hand Articulations using Kinect," The 22nd British Machine Vision Conference, pp.101.1-101.11, 2011.9.
- [9] J. Lee, K. Oh, "Yoga Learning Content by using Kinect", HCI 2012, pp.221-223, 2012.1.