

모션기반 터치리스 인터랙티브 증강현실 어플리케이션

양기선, *김창현

KBS 기술연구소, *고려대학교

ksyang@kbs.co.kr, *jhkim@korea.ac.kr

Touchless Motion Based Interactive Augmented Reality Application

Ki-Sun Yang *Chang-Hun Kim

KBS Technical Research Institute *Korea University

요 약

본 논문은 방송 콘텐츠에서 사용할 수 있는 터치리스 형태의 모션기반 인터랙티브 증강현실 어플리케이션을 제안한다. SD 급 해상도를 가진 범용 모션 카메라를 사용하여 HD 해상도의 방송에서 증강현실 어플리케이션을 구현하기 위해서 깊이 영상의 임의 범위를 지정하여, HD 해상도에 매칭시켜 2D 터치리스 영역에서 손의 멀티터치를 인식하게 하였다. 이렇게 인식된 모션을 자주 사용하는 ‘왼손 클릭’, ‘오른손 클릭’, ‘양손 클릭’, ‘스크러빙-상/하/좌/우’ 제스처를 인식하게 하였으며, ‘카루젤 메뉴’, ‘막대그래프 메뉴’, ‘슬라이드 메뉴’, ‘큐브 메뉴’ 를 구현하여 터치리스 형태의 모션기반 인터랙티브 증강현실 어플리케이션을 제안한다. 방송에서 자주 사용하는 그래픽 정보를 연기자나 사회자가 본 논문에서 제안하는 인터랙션 어플리케이션을 통하여 오퍼레이터 없이 스스로 더 능동적으로 정보를 표출할 수 있는 방법을 설명한다.

1. 서론

몇 년 전부터 스마트 폰이 많이 배포 됨에 따라 터치 인터페이스를 사용하기 위한 ‘ 터치 제스처’ 가 우리에게 많이 익숙해져 있다. 터치를 위해서는 터치 스크린이라는 하드웨어 인터페이스가 필요하며, 터치 영역 또한 터치 스크린으로 제한된다. 최근 방송에서도 터치 스크린을 이용한 방송 제작을 자주 사용하고 있는 추세에 있다.

최근, PrimeSense 사[1]의 모션인식 카메라, Microsoft XBOX Kinect[2]라는 가정용 콘솔 게임의 인터페이스가 출시되어, 사용자들로 하여금 전통적인 마우스, 키보드 인터페이스 형태가 아닌 동작(제스처)을 통하여 컴퓨터와의 인터랙션을 가능하게 하고 있어 큰 호평을 받고 있다. 그 장치는 사용자에게 모션인식 카메라로부터의 깊이 영상과 RGB 영상을 제공하며, 연구원이나 개발자가 깊이 영상 자체를 활용하거나 또는 그 영상의 분석을 통하여 추출된 휴먼 골격추적정보를 사용하여 사용자의 동작을 응용할 수 있게 하고 있다. 근래에 이런, Kinect 를 사용하여 게임 인터페이스 뿐 만 아니라 그 외에도 다양한 응용사례가 보고되고 있다.

예를 들어, 먼저, 의사가 수술 중에 모니터 화면 제어를 위한 인터페이스로 사용하거나[3], 가상의 캐릭터를 제어하는 아바타 제어용 인터페이스로 사용하거나[4], 사용자의 위치와 거리를 추적하여 다른 시점의 영상을 보여주거나, 손의 모션을 통해 글을 쓰는 인터페이스로 사용되는 것처럼[5] 사용자의 골격 추적 정보를 응용하여 다양한 인터페이스로 활용하는 사례가 보고되고 있으며, 또, 깊이 영상 자체를 활용하여 실사 3D 물체를 그래픽 데이터로 모델링하는 3D 복원에 사용되기도 하고[6], 휴먼 페이스 복원을 통하여 실시간 페이스 애니메이션에 활용된 사례도 있다.

본 논문은 방송에서 자주 사용하는 그래픽 정보를 연기자의 모션을 통하여 방송 콘텐츠를 제작할 수 있도록 하기 위해서, 터치리스 형태의 모션기반 인터랙티브 증강현실 시스템을 제안한다. 기존의 인터랙티브 증강현실 시스템은 특정 마커를 사용하여 그래픽과의 인터랙션을 시도하곤 하였다. 그러나, 방송에서 마커의 사용은 마커를 인식 할 때 조명과 같은 외부 환경에 마커 인식이 매우 민감하여 영상처리의 부정확함으로 인한 그래픽 데이터의 떨림 현상이 나타나거나, 마커인식 실패로 인하여 그래픽이 사라지거나 튀는 현상으로 인하여 실시간으로 제작하는 방송 현장에서는 활용하기가 어려운 점이 있다. 최근 MS 와 PrimeSense 사는 Kinect 와 같은 깊이 영상 추출 센서 기반의 자사의 카메라를 활용할 수 있는 SDK 를 공개하였다. 이 SDK 는 특히 연기자의 모션 추적이 강인하고, 마커가 필요하지 않는 등 진행자와 인터랙티브한 증강현실 시스템을 구현하기에 적합한 인터페이스로 활용될 수 있음을 확인할 수 있었다.

본 논문은, 물리적인 하드웨어 터치 기반 인터페이스가 아닌, 이런 범용 깊이센서 카메라를 이용하여 터치리스 기반의 인터랙티브 증강현실 응용 시스템을 제안하고자 한다. 사용자의 특정 동작, 예를 들어, 즉, 스크러빙 모션(상,하,좌,우), 클릭 다운/업 모션을 인식하여 다양한 비주얼한 그래픽 방송 데이터 정보를 제작할 수 있도록 사용자의 모션에 따라 제어되는 인터랙티브 증강현실 방송 시스템을 구현하였다.

본 논문의 장점은 다음과 같다.

- 방송에서 연기자에 진행자의 자유도를 높일 수 있는 터치리스 인터랙티브 증강현실 시스템을 제안한다.
- 마커가 필요 없이 그래픽 정보를 진행자의 모션에 따라 다양한 스토리로 제작할 수 있다.

- SD급 Kinect 인터페이스 기기와 방송용 HD 카메라간의 간단한 그래픽 정합방법을 보여준다..

- 방송에서 자주 사용하는 형태의 정보 그래픽을 사용하여, 연기자의 모션을 통하여, 큐브 인터랙션, 스크리빙 인터랙션, 막대그래픽 인터랙션, 카루젤 인터랙션으로 구현된 방송 콘텐츠를 제안한다.

2. 시스템 구성

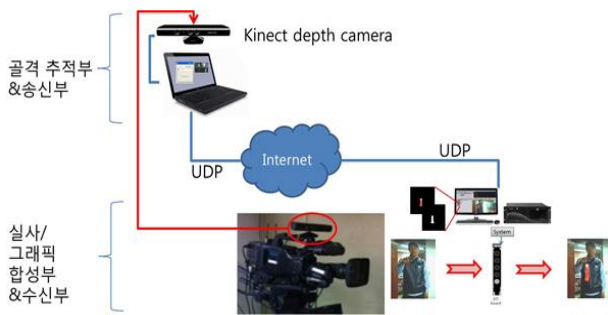


그림 1. 시스템 구성

그림 1은 전체 시스템 구성을 보여준다. 이 시스템은 크게 두 부분으로 구성 되어 있다. 하나는 골격을 추적하여 UDP로 추적된 정보를 전송하는 송신부와 다른 하나는 이 정보를 받아서 방송용 카메라 영상과 실시간으로 그래픽을 합성하는 수신부로 구분된다.

3. 터치리스 모션인식 방법 및 모션 정의

그림 2 송신부는 터치리스 형태의 제한된 영역에서의 모션을 인식한 그림을 보여 주고 있다.

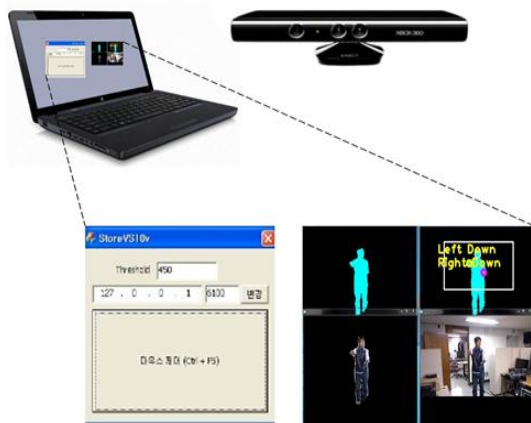


그림 2. 송신부 UI

Sender 부의 시스템은 노트북과 모션인식 카메라로 구성된다. OpenNI[7], OpenCV[8]를 사용하여 구현하였으며, 모션 추적을 위해서는 먼저 다음과 같은 설정을 해야 한다.

- 거리값 설정 : 어깨와 손까지의 거리 값
- 서버 설정 : 전송 서버 IP

프로그램을 실행하면, 하나는 RGB 영상, 다른 하나는 깊이 영상이 보여진다. 가장 먼저 수행해야 하는 단계는 방송용 카메라와 모션인식 카메라의 영상간의 제한된 동작 영역을 선택하는 것이다. 이것이 필요한 이유는 추적은 SD급 해상도의 모션인식 카메라를 통해서 수행되고 합성은 방송용 HD 카메라 영상 위에 수행되기 때문이다. 터치 스크린영역을 구성하기 위해, 또, 구현의 간편함을 위해서 모션인식 카메라의 깊이 영상과 방송용 카메라의 영상은 서로 다른 화각, 해상도를 가지고 있으므로 방송용 카메라의 줌을 이용하여 적절한 화각을 조절한다, 모션인식 카메라를 방송용 카메라 위에 위치시켰다. 두 카메라간의 보정을 위한 복잡한 계산을 하지 않고, 두 카메라의 화각이 겹치는 영역을 수동으로 선택하게 하였다. 방송용 카메라의 화각은 줌 컨트롤을 통하여 조절한다. 조절된 HD 카메라 화각의 오른쪽 끝과 왼쪽 끝이 깊이 영상 내의 위치하도록 한다. 이렇게 함으로써, 수동으로 설정된 일정 거리에서만은 두 영상의 2D 위치가 비슷하게 되어, 모션인식 카메라의 깊이 영상으로부터 추적된 골격 포인트들은 HD 방송 카메라영상에서는 위치와 매우 유사하게 된다. 기존의 마커 기반의 증강현실 시스템들에서는 특징점 추적과 같은 것을 위해서 반드시 정확하게 위치를 추적해야 했지만, 모션 인식 카메라로부터 추적된 골격 포인트는 기존의 증강현실 어플리케이션의 특징점처럼 사용하는 것은 적절치 않다. 따라서, 이렇게 제약된 거리에서 결정된 2D 영역에서 모션인식 카메라의 응용은 모션의 대략적인 추적과 동작의 인지만으로도 충분히 인터랙티브한 합성 결과물을 얻을 수 있다.

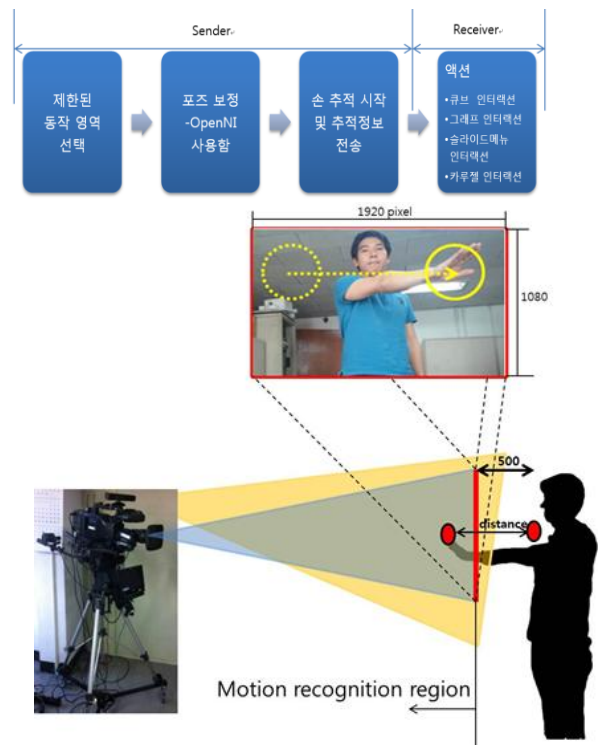


그림 3. 동작흐름도 및 동작 이벤트 처리

사용자는 그림 3 과 같은 과정을 따른다. 먼저, 모션인식

카메라 RGB의 해상도와 방송용 카메라의 해상도가 서로 다르기 때문에 방송용 카메라의 줌/인 인터페이스를 이용하여 모션인식 카메라의 화각에 포함 되도록 조절한다. 모션인식 카메라를 방송용 카메라 위에 위치 시킴으로써 서로 유사한 뷰를 갖게 수동으로 조절한다. 이렇게 조절을 시킨 후, 모션인식 카메라의 RGB 영상 위에 마우스를 이용하여 제한된 영역을 선택하여 이 영역을 '인터랙션 영역'으로 제한한다. 제한된 인터랙션 영역이 선택이 되었으면, OpenNI를 이용하여 포즈 보정과정을 수행하여 골격 추적 초기화를 수행한다. 보정이 완료되자마자 사람의 얼굴, 양손의 특징점 3D 정보를 서버로 UDP 형태로 전송하게 된다. 모션의 인식 순서는 클릭 다운, 무브, 클릭 업과 같이 세 부분으로 나뉘어진다.

클릭 다운은 모션 인식이 시작되는 부분이다. 사용자는 일정 거리(기본값 500, 약 0.5m)내에서는 자유롭게 손을 움직일 수 있고, 만약, 그 손이 미리 정해 놓은 거리(500)를 넘어서면, 이후 프레임부터는 사용자의 모션을 인식하는 준비를 하게 된다. 따라서, 클릭 다운은 모션 인식의 시작이 된다. 그 다음은 실제로 동작을 취하는 무브 단계이다. 무브 단계에 들어서면, 모션을 인식하게 되는데, 본 논문의 시스템 구조에서 가장 빈번하게 사용될 모션으로 클릭 다운/업, 스크러빙(좌우상하) 모션을 인식하여 인터랙티브 방송 콘텐츠(큐브/막대그래프/슬라이드/카루젤 인터랙션)에 연개하였다. 무브 단계가 끝나고 다시 사용자가 자유로운 동작을 취할 수 있도록 무브 동작의 끝으로 클릭업 단계에 들어선다. 클릭 업은 모션 인식이 끝나는 부분, 즉, 사용자의 모션 인식이 종료되는 부분으로 이후 일정 거리 내에서 자기 자유롭게 손을 움직이더라도 어떤 처리도 하지 않는다.

서버에서는 VizRT사의 VizArtist 툴을 사용하여 그래픽을 합성하였고, Matrox X.Mio 보드를 통하여 방송신호를 출력하였다. VizArtist는 일반적으로 방송에서 많이 사용되는 3차원 가상스튜디오용 그래픽 제작 도구로서, 이 툴에서는 Visual Basic Script를 사용하여 외부 포트로부터 들어오는 데이터를 수신할 수 있고, 그 정보는 툴에서 제공하는 그래픽 오브젝트의 움직임을 제어 하기 위해 사용될 수 있다. 서버 시스템에서는 Sender에서 보내온 골격 정보로부터 사용자의 양손의 상,하,좌,우 스크러빙 모션을 인식하여 그래픽 오브젝트의 애니메이션을 제어하기 위해 사용되었다.



그림 4. 좌우 인터랙션 위한 로고메뉴와 큐브 메인메뉴

콘텐츠의 시작은 그림 4에서 보는 것과 같이 KBS 로고와 큐브 메뉴로 구성되었다. KBS 로고와 큐브 메뉴는 두 가지 모두 특정 방송 콘텐츠의 시작 메뉴로 사용될 수 있다.

표 1. 실험 변수 설정

모션 정의	모션 표시
클릭	
클릭 오른쪽 스크러빙	
클릭 왼쪽 스크러빙	
클릭 위 스크러빙	
클릭 아래 스크러빙	
클릭 좌우 스크러빙	

표 1은 제안된 모션을 보여준다.

큐브 메뉴는 아래, 위 면을 제외한 네 면에 각기 다른 메뉴를 구성하였다. 큐브 메뉴 인터랙션은 큐브가 특정 응용 메뉴 시작의 사용자 인터페이스가 되도록 하였다. 큐브에는 특정 메뉴를 구성하여, 큐브를 돌림으로써 원하는 인터랙션 메뉴가 선택되도록 하였다.

막대 그래프 메뉴 인터랙션은 사용자가 직접 막대 그래프를 손의 모션을 통하여 제어 하게 하였다. 오른손의 클릭 다운 후 무브 모션을 통하여 원하는 높이 만큼 위치한 후 클릭 업 모션을 통하여 놓으면 막대 그래프의 길이가 사용자의 손의 위치를 따라 길어지거나 짧아진다.

스크러빙 메뉴 인터랙션은 오른손의 스크러빙 모션을 통하여 그래픽 오브젝트의 애니메이션을 제어함. 원하는 오브젝트가 정면에 왔을때, 왼손을 업 스크러빙하면서 그 선택된 오브젝트가 팝업되고, 다운 스크러빙 함으로써 원래의 위치로 돌아가게 된다.

카루젤 메뉴 인터랙션은 사용자가 오른손으로 클릭 다운 모션을 취할 경우, 원형의 카루젤 메뉴가 손에 위치에 따라 다니게 된다. 왼손으로 스크러빙 모션을 통하여 원하는 그래픽 오브젝트를 선택되도록 돌릴 수 있으며, 본 시스템에서는 카루젤 메뉴로부터 선택된 메뉴를 지도 그래픽 오브젝트와 연동하는 예제를 구현하였다.

4. 실험 결과

표 2. 실험 변수 설정

표 2는 제안된 손의 모션, 클릭, 좌/우, 상/하 클릭 스크러빙으로 큐브/막대그래프 1,2/슬라이드/카루젤 메뉴를 통해서 인터랙티브 방송 콘텐츠 결과 영상을 보여주고 있다.

5. 결론

본 논문에서는 SD급 해상도의 모션인식 카메라를 이용하여 HD 방송환경에서 터치리스 형태의 연기자의 모션을 이용한 인터랙티브 증강현실 어플리케이션을 제안하였다. 제안된 방법으로 만들어진 콘텐츠는 선거방송이나 뉴스,

음악차트 프로그램등 다양한 방송 프로그램 제작 현장에서 그래픽 방송 정보를 연기자의 모션을 통하여 인터랙티브한 방송 콘텐츠 제작이 가능하다. 향후, 연기자가 가상스튜디오안에서 3 차원 공간에서도 인터랙티브하게 그래픽을 연동하는 방송 제작에 적용할 예정이다.

참고 문헌

- [1] <http://www.primesense.com/>
- [2] <http://en.wikipedia.org/wiki/Kinect>
- [3] <http://www.youtube.com/watch?v=f5Ep3oqicVU>
- [4] http://www.youtube.com/watch?v=zU1_HB4sMc0&list=PLD7A6D0907465FB85
- [5] <http://www.youtube.com/watch?v=cYQ07SP7z8w>
- [6] http://www.youtube.com/watch?v=of6d7C_ZWw
- [7] 'OpenCV', <http://opencv.org/>
- [8] 'OpenNI', <http://www.openni.org/>

모션기반 인터랙티브 콘텐츠 실험 영상	
큐브	
막대 1	
막대 2	
슬라이드	
카루젤	