

Depth 기반 3D 증감 방송을 위한 객체 인지 기술

추현곤, *김민영, *조용주, 최진수, 김진웅

한국전자통신연구원, *상명대학교

hyongonchoo@etri.re.kr

Object Registration for Depth-based Augmented 3D Broadcasting Service

Hyon-Gon Choo, *Minyoung Kim *YoungJu Cho, Jin Soo Choi, and Jinwoong Kim

ETRI, *Sangmyong University

요약

최근 3D 입체 영화 및 3D 방송 기술의 개발과 더불어, 사용자의 몰입감을 극대화하기 위하여, 3D 입체영상과 사용자의 인터랙션 효과를 융합하는 3D 증감 방송 기술에 대한 요구가 높아지고 있다. 본 논문에서는 3D 증감방송을 위해, 3DTV 환경에서의 사용자 인터랙션 서비스에 대한 시나리오에 대해서 분석하고, 이를 바탕으로 제한된 시나리오를 위한 객체 인지 및 제어 기술에 대해서 소개한다. 본 논문에서 깊이 정보와 연계된 3D방송 환경을 가정하고, 이를 TV 환경에 적용했을 경우의 인터랙션 시나리오에 제시하였으며, 다 수의 깊이 카메라를 이용하여 시나리오에 필요한 객체 인지 방법에 대해서 제시한다. 키넥트 카메라를 이용한 구현을 통해 객체 인지 및 시점 전환 구현 결과를 보여준다.

1. 서론

3D 입체 영화 및 3D 방송 기술의 개발과 더불어, 사용자의 몰입감을 극대화하기 위하여, 3D 입체영상과 사용자의 인터랙션 효과를 융합하는 3D 증감 방송 기술에 대한 요구가 높아지고 있다 [1]. 최근 개발되는 TV의 경우, 얼굴 인식과 같은 사용자 인식 및 제스처 인식 기능을 도입하여 사용자와 인터랙션을 제공할 수 있는 환경을 제공하고 있다. Fantastic IPTV Remote Control에 대한 연구[2]에서는 자이로 센서와 3차원 가속도 센서를 사용해서 2차원 마우스와 같은 역할을 할 뿐만 아니라 제스처 동작도 지원 및 게임용 컨트롤러로서도 사용 가능한 사용자 인터페이스를 제시하고 있으며, MIT 미디어 랩의 DepthJS[15]는 키넥트(Kinect)와 컴퓨터 비전 기술, 그리고 자바스크립트 언어를 활용해서 영화 마이너리티 리포트의 제스처 인터페이스를 구현하였다.

키넥트 카메라의 개발과 더불어서 최근 3D환경과 연계된 증강현실 기술에 대한 연구도 활발하게 이뤄지고 있다. 미국 텍사스 대학의 Xia 등은 2차원 머리 컨투어 모델 (2D Head Contour Model)과 3차원 머리 표면 모델 (3D Head Surface Model)을 사용하여 사람을 탐지하고 주변환경으로부터 사람을 구분해내는 방법[4]을 제안하였으며, Saenko 등은 키넥트(Kinect)센서를 이용해서 객체와 객체의 종류를 탐지할 수 있는 방법을 제안하였다 [5]. Mohd Kufaisal 등은 자연적 사용자 인터페이스 (Natural User Interface)에 관한 연구에서 키넥트 (Kinect)같은 깊이 카메라가 사용자 인터페이스에 대한 정확성과 효율성을 높이기 위해 활용될 수 있음을 언급하였다 [6].

본 논문에서는 3D 증감방송을 위해, 3DTV 환경에서의 사용자 인터랙션 서비스에 대한 시나리오에 대해서 분석하고, 이를 바탕으로 사용자 인터랙션 서비스 시나리오를 위한 객체 인지 및 제어 기술에 대

한 구현 결과에 대해 소개한다.

본 논문의 구성은 다음과 같다. 2 절에서는 3D 증감방송을 위한 사용자 인터랙션 서비스 시나리오에 대해 살펴본 후, 3 절에서는 이를 위한 깊이 기반 객체 인지 방법에 대해서 살펴본다. 4 절에서는 논문에서 제안하는 기법의 구현 결과를 설명하고, 6절에서 결론을 맺는다.

2. 3DTV 환경을 위한 사용자 인터랙션 서비스 시나리오

본 논문에서 3D 증감방송을 3DTV 환경에서의 적용하기 위한 사용자 인터랙션 시나리오에 대해서 분석하였다. 사용자 인터랙션 시나리오의 분석에 있어, 3D 방송은 현재의 양안식 3D 뿐만 아니라 다시점, 홀로그래픽 3D등을 고려하여 여러시점의 화면과 깊이 정보가 존재하는 것을 가정하였으며, 이러한 기술을 활용해서 영상에서 객체 인식을 좀 더 쉽게 하고, 그 기술을 바탕으로 새로운 형태의 3D TV 환경에 적용시켰을 때의 인터랙션 서비스 시나리오에 대해 분석하였다.

분석한 결과를 바탕으로 한 3D 증감 방송을 위한 사용자 인터랙션을 표 1과 같다.

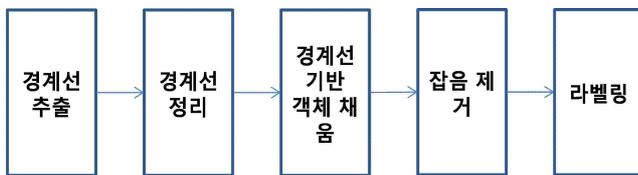
<표 1> 3D증감방송을 위한 사용자 인터랙션 시나리오

1	3차원 TV 공간에서의 이동 시나리오
2	3차원 TV 공간에서의 객체 선택 시나리오
3	3차원 TV 공간에서의 객체 조작 시나리오
4	3차원 TV의 광고 시나리오
5	시점 변환 시나리오
6	객체에 대해 태깅 시나리오
7	입체 영상 방송과 SNS의 연계 시나리오
8	객체 단위의 스크랩(Scrap) 시나리오

3. 깊이 정보 기반 3차원 데이터를 이용한 객체 인지

3D 증감 방송의 사용자 인터랙션을 위해서는 3차원 정보를 통해, 사용자가 반응하고자 하는 객체를 인지하는 작업이 매우 중요하다. 본 논문에서는 2절에서 제시된 사용자 인터랙션 시나리오 중 객체 선택/태깅/시점 변환 시나리오 등에 적용이 가능한 객체 인식 방법에 대해서 제시한다. 서로 다른 시점에서 획득된 깊이 정보와 시점 영상을 이용하여, 각 시점에서 객체를 분리하고 분리된 물체의 위치 좌표 관계를 이용하여 객체를 인지하고 이와 연계된 정보(태그)를 출력하거나 시점을 변화한 정보를 확인할 수 있다.

본 논문에서 객체 인식 방법은 다음과 같이 이뤄진다. 프레임 별로 입력된 깊이 영상과 시점 영상에 대해서 <그림 1>과 같은 전처리 작업을 거쳐서 각각의 객체를 분리한다.



<그림 1> 객체 인식을 위한 전처리 과정

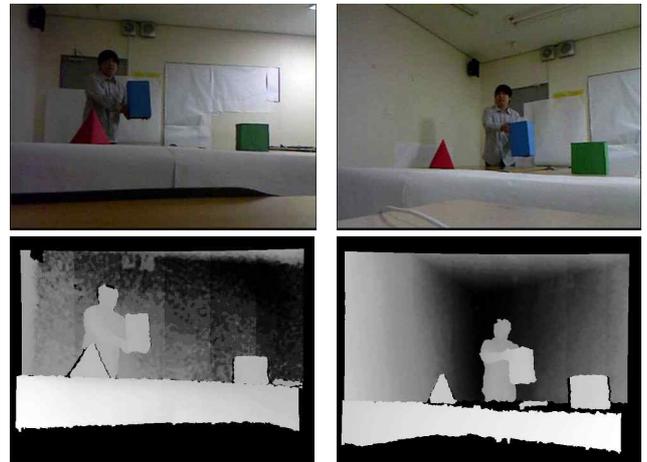
깊이 영상과 시점 영상의 경계선 추출을 위해서는 Mean-shift 알고리즘을 이용하여 단순화한 영상에 대해 Canny Edge detector를 사용하였으며, 깊이 영상과 시점 영상의 경계선 일치 여부를 통해 경계선을 정리하였다. 생성된 경계선을 기반으로 Hole filling을 이용하여 객체를 분리하는 작업을 거친 후, 가로 및 세로 방향으로 퍼져 나오는 노이즈를 제거한 후, Blob 기반 라벨링 알고리즘을 사용하여 객체를 분리하였다.

분리된 객체에 대하여 초기 객체에 대한 깊이 정보 및 태깅을 위한 메타데이터 입력을 통해 객체를 등록하도록 하였으며, 마지막으로 객체 인식이 되는 영상 이미지들을 다시 합쳐서 동영상의 형태로 만들고, 사용자의 상호작용을 통해 이름을 붙여서 보여주거나 시점 변환을 지원할 수 있도록 개발하였다

3. 구현 결과

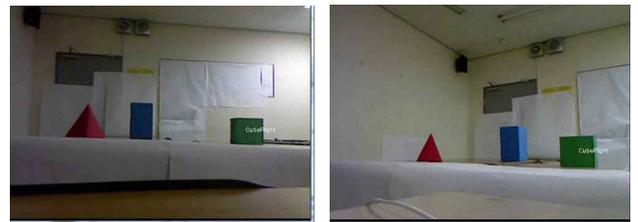
두 대의 Kinect 카메라를 활용해서 서로 다른 시점에서의 RGB 컬러 영상과 깊이 영상을 촬영하였다. 제한된 공간에서 사각뿔, 정육면체, 직육면체 같은 물체들을 놓고 촬영하였다. 두 대의 카메라는 3차원 좌표 구성의 편의성 및 간섭을 고려하여 직교하도록 배치되었으며, 촬영된 영상을 기반으로 사용자가 화면을 보면서 반응에 따라 결과를 보여주도록 플레이어를 구현하였다. <그림 2>는 실험을 위해 촬영된 영상에 대한 예를 보여준다.

객체 인식에 필요한 이름은 메타데이터를 통해 제공하도록 하였다. 구현결과에서 촬영된 결과를 서로 다른 시점에서 볼 수 있으며, 사용자가 객체 선택하면 인식된 객체에 대한 정보를 출력하도록 하였다.



<그림 2> 객체 인지 실험 환경

시점 변환이 일어난 경우, 이전에 사용자가 선택되었던 객체를 추적하면서 정보를 보여줄 수 있도록 만들었으며, 객체가 화면상에서 움직이더라도 그 객체를 따라다니면서 이름을 보여줄 수 있도록 개발하였다. <그림 3>와 <그림 4>는 개발한 플레이어를 이용한 인터랙션에 대한 결과를 보여주고 있다.



<그림 3> 객체 이름 태그 출력과 시점 변환



<그림 4> 사용자 인터랙션에 따른 태그 출력

4. 결론

3차원 증감 방송은 단순하게 입체 영상을 보여주는 것이 아니라 시청자로 하여금 방송 안의 객체들과 다양한 상호작용을 할 수 있도록 해야한다. 본 논문에서는 3D 증감방송을 위해, 3DTV 환경에서의 사용자 인터랙션 서비스에 대한 시나리오에 대해서 분석하였다. 본 논문에서 향후 깊이 정보와 연계된 3D방송 환경을 가정한 여러 가지 사용자 인터랙션 시나리오를 제시하였으며, 키넥트 카메라를 이용하여 시점 전환 및 태깅과 같은 시나리오에 필요한 객체 인지 방법에 소개하였다.

추후에는 객체 인식 기술을 기반으로 한 다양한 사용자 인터랙션 서비스가 적용된 3D 증감 방송에 대한 연구를 진행될 예정이다.

Acknowledgement

본 연구는 방송통신위원회 및 한국방송통신진흥원의 방송통신미디어 원천기술개발 과제의 일환으로 수행되었음. [과제번호 10912-02001, 지상파 양안식 3DTV 방송시스템기술개발 및 표준화]

참고문헌

- [1] Hand, C. "A Survey of 3D Interaction Techniques," Computer Graphics Forum, Vol. 16, pp. 269-281, 2003
- [2] Fantastic IP TV Remote Control, <http://www.youtube.com/watch?v=pRMpbe36TKw>.
- [3] DepthJS, <http://depthjs.media.mit.edu/>
- [4] L. Xia, C.-C. Chen, and J. K. Aggarwal, "Human Detection Using Depth Information by Kinect", International Workshop on Human Activity Understanding from 3D Data in conjunction with CVPR (HAU3D), Colorado Springs, CO, June 2011.
- [5] K. Saenko, S. Karayev, Y. Yia, A. Shyr, A. Janoch, J. Long, M. Fritz and T. Darrell, "Practical 3-D Object Detection Using Category and Instance-level Appearance Models," in Proceedings of International Conference on Intelligent Robots and Systems, 2011.
- [6] Mohd Kufaisal bin Mohd Sidik, Mohd Shahrizal bin Sunar, Ismahafezi bin Ismail, Mohd Khalid bin Mokhtar, and Normal binti Mat Jusoh, "A Study on Natural Interaction for Human Body Motion Using Depth Image Data," in Proceedings of Workshop on Digital Media and Digital Content Management, 2011.