

다시점 카메라 및 depth 카메라를 이용한 3 차원 비디오 생성 기술 연구

엄기문*, 장은영**, 허남호***, 이수인****

한국전자통신연구원 디지털방송연구단 방송시스템 연구그룹

A Study on the 3D Video Generation Technique using Multi-view and Depth Camera

Gi-Mun Um*, Eunyoung Chang**, Namho Hur***, and SooIn Lee****

*Digital Broadcasting Research Division, Electronics and Telecommunications Research
Institute,

E-mail : *gmum@etri.re.kr, **eychang@etri.re.kr, ***namho@etri.re.kr, ****silee@etri.re.kr

Abstract

This paper presents a 3D video content generation technique and system that uses the multi-view images and the depth map. The proposed uses 3-view video and depth inputs from the 3-view video camera and depth camera for the 3D video content production. Each camera is calibrated using Tsai's calibration method, and its parameters are used to rectify multi-view images for the multi-view stereo matching. The depth and disparity maps for the center-view are obtained from both the depth camera and the multi-view stereo matching technique. These two maps are fused to obtain more reliable depth map. Obtained depth map is not only used to insert a virtual object to the scene based on the depth key, but is also used to synthesize virtual viewpoint images. Some preliminary test results are given to show the functionality of the proposed technique.

I. 서론

최근 3 차원 비디오에 대한 관심이 컴퓨터 비전이나 컴퓨터 그래픽스 분야에서 높아지고 있다. 3 차원 비디오는 여러 가지 형태의 비디오를 포함하고 있는데, 옴니 비디오, 자유시점 비디오, 그리고, 다시점 비디오

등이 그것이다. MPEG-4 3DAV ad-hoc 그룹에서는 다양한 형태의 다시점 비디오를 어떻게 인코딩(encoding)하고, 전송하고, 디코딩(decoding)하여 재생할 것인가에 대한 표준화가 진행되고 있다[1].

한편, 디지털 방송 분야에 있어, 많은 전문가들은 3DTV 가 HDTV 이후 차세대 디지털 TV 형태 중의 하나가 될 것이라고, 예측하고 있다. 2D 비디오와 비교하여 3D 비디오는 다양하고 새로운 부가 서비스를 시청자 또는 제작자에게 제공한다. 이러한 3DTV 의 궁극적인 목표는 자유시점 TV 로서, 이를 통해 사용자는 동시에 여러 시점에서 촬영된 장면을 인터랙티브(interactive) 하게 선택하여 임의의 시점에서 시청 가능하게 된다. 그러나, 시점의 개수와 카메라 대수 간에는 trade-off 가 있어서 보다 많은 시점을 제공하기 위해서는 더 많은 대수의 카메라가 필요하게 된다. 또한 최근의 3 차원 비디오는 가상 현실 또는 혼합 현실(virtual and mixed reality) 응용과 결합되어 많이 이용되고 있다[2]. 많은 수의 연구자들이 완전 3 차원 콘텐츠를 생성, 전송, 디스플레이 하기 위한 방법에 대해 연구를 수행하여 왔다. 그러나, 완전 3 차원 콘텐츠의 실시간 제작은 계산속도와 아직 그래픽 기술의 발전에도 불구하고 아직 불가능하다. 따라서, 대부분의 3 차원 콘텐츠를 이용한 응용이 TV 나 영화 프로그램의 사후제작(post-production) 등으로

본 논문은 정보통신부의 연구사업인 “지능형 방송 통합 정보 방송(SmarTV) 기술 개발” 과제의 일환으로 수행된 결과입니다.

제한되고 있다.

본 논문에서는 3 시점 카메라와 depth 카메라를 이용하여 3 차원 비디오 콘텐츠를 생성하는 시스템을 제안한다. 제안한 시스템은 O. Grau 등이 제안한 3D video 제작 시스템[3] 또는 영상기반 가상 스튜디오(image-based virtual studio)[2]와 유사하나, 제안한 시스템은 청색 배경의 크로마 키(chroma key) 대신, depth 카메라를 이용한 depth key 를 사용하여 배우와 배경을 분리하였다는 측면에서 기존의 시스템과 차별된다.

본 논문의 구성은 II 장에서 3 차원 비디오 콘텐츠 생성시스템들에 대해 알아보고, III 장에서는 본 논문에서 제안한 3 차원 비디오 콘텐츠 생성 시스템에 대해 살펴보고, IV 장에서는 제안한 3 차원 비디오 콘텐츠 생성 시스템에 의해 3 차원 비디오 콘텐츠를 생성한 결과에 대해 간략히 소개하며, V 장에서 결론을 맺는다.

II. 기존의 3 차원 비디오 콘텐츠 생성 연구

최근 수 년 동안, 3 차원 비디오 콘텐츠 생성에 대해 많은 연구가 이뤄져 왔으며, 이러한 연구는 주로 장면의 구성요소 즉, 배우 및 배경 등을 3 차원으로 모델링 하기 위한 기술에 대해 이루어졌다. O. Grau 등 [3]은 다시점 카메라와 크로마 키(chroma key) 장치를 이용하는 3 차원 모양 복원 및 렌더링 시스템을 구현하였다. 이 때, 배우의 3 차원 모델링을 위해 주로 shape-from-silhouette 을 사용하였다. 또한 배경으로부터의 배우 분리를 위해 주로 크로마 키 또는 retro-reflective cloth 와 청색 LED 링을 조합하여 사용하였다. 한편, 배경 장면의 모델링을 위해서는 별도의 배경 장면 모델링 기법을 사용하였다. 이외에 K. Tomiyama 등[4]은 19 대의 fire-wire 카메라를 사용하여 동적인 3 차원 객체 생성 시스템을 개발하였다. 그들은 움직이는 인체를 volume intersection 기법 및 스테레오 정합 기법을 결합하여 모델링 하였다. 이 연구에서도 Grau 등의 연구결과와 마찬가지로, 배우 실루엣의 추출을 위해 청색 스크린(blue screen)이 사용되었다. 박종일 등[2]은 실사영상기반 비디오 합성 시스템을 개발하였다. 이 시스템에서는 가운데 하나의 카메라를 기준으로 다른 4 개의 카메라를 수평과 수직 방향으로 대칭이 되도록 배치하였으며, 다시점 스테레오 정합 기법을 통하여 중심 카메라에 대한 깊이 맵(depth map)을 생성하였다. 이 깊이 맵은 중간

입의 시점 생성과 다른 장소에서 촬영된 배우 영상을 합성하는데 depth key 로서 이용되었다. 그러나, 별도 촬영 영상에서 배우의 분리에서는 여전히 크로마 키를 사용하였다.

이러한 기존 연구에 비해 본 논문에서는 다시점 및 depth 카메라를 사용하여 중간 가상 시점 영상 생성 및 깊이 기반 3 차원 장면 합성을 지원하는 3 차원 비디오 콘텐츠 생성 시스템을 제안한다.

III. 제안한 3 차원 비디오 콘텐츠 생성 시스템

본 절에서는 제안한 3 차원 비디오 생성 시스템의 주요 특징에 대해 설명한다. 그림 1 은 본 논문에서 제안한 3 차원 비디오 콘텐츠 생성 시스템의 구조 및 기능을 나타낸 것이다. 제안한 시스템은 크게 3 시점 영상 및 깊이 맵 획득 및 전처리, 3 시점 카메라 보정, 변이/깊이 맵 추정, 움직임 객체 추출 및 추적, 중간 가상 시점 영상 합성, CG 객체/배경 카메라 파라미터, 3D 장면 합성(배경 대치, CG 객체 삽입) 등의 기능을 가지고 있다. 각 기능별로 세부적으로 설명하면 다음과 같다.

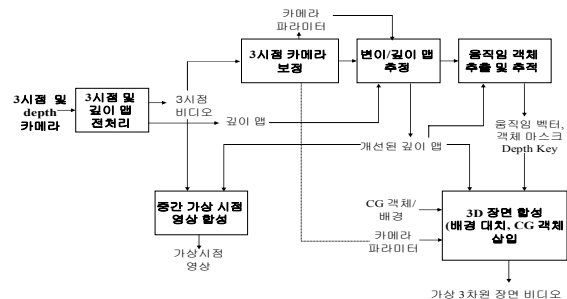


그림 1. 제안한 3 차원 비디오 콘텐츠 생성 시스템

3.1.3 시점 영상 및 깊이 맵 획득 및 전처리

본 연구에서는 3 시점 영상 획득 및 가운데 시점에 대한 깊이 맵 획득을 위하여 그림 2 와 같은 720 X 486 해상도를 가지는 방송용 3 시점 비디오 카메라 및 depth 카메라를 사용하였다. Depth 카메라로서는 3DV Systems 사의 ZCam™ 을 사용하였다. 표 1 은 ZCam™ 카메라의 출력은 4 채널 영상 캡처장치(frame grabber)를 이용하여 동기를 맞추어 획득되었다. Depth 카메라에 의한 깊이 맵의 깊이 범위는 장면에 따라 조절될 수 있다.



그림 2. 3 시점 및 depth 카메라 시스템

표 1. ZCam™ depth 카메라 사양 [5]

깊이 범위	0.5 to 7 m (1.5 - 23 feet)
깊이 해상도	1cm(설정깊이범위 의존)
FOV(시야각)	40 도
출력 포맷	NTSC or PAL (4:3, SD)
출력 신호	SMPTE 259 M (D1)

이렇게 획득된 3 시점 영상 및 깊이 맵은 변이 추정
 정의 정확도 향상을 위하여 각 시점 영상 간에 컬러
 를 일치시키기 위한 컬러 보정을 수행하며, 깊이 맵
 에 대해서는 잡음 제거를 위하여 시간 축 및 공간 축
 상에서 중간 값 필터링(median filtering)을 수행하였다.

3.2.3 시점 영상으로부터의 변이 추정

최근 스테레오 정합 기법의 경향은 두 개 이상의
 깊이/변이 추정 기법을 사용하여 하나의 기법만을 사
 용하였을 때의 단점을 다른 기법을 동시에 사용함으
 로써 보완하는 기법들이 주류를 이루고 있다[6]. 본
 연구에서 사용한 depth 카메라에 의한 깊이 맵은 장
 면 내 전체 깊이 범위에 대한 깊이 맵을 얻을 경우
 물체와 물체간, 물체와 배경 간의 상대적 깊이는 잘
 표현하지만, 물체 내부의 정밀한 깊이 변화는 표현하
 지 못한다. 반면에 다시점 스테레오 정합 기법은 차
 페 영역이나 텍스처가 없는 영역에서 오정합을 많이
 발생한다. 따라서, 본 연구에서는 먼저 depth 카메라
 로부터 얻어진 깊이 맵을 3 시점 영상에 대한 변이
 추정시 탐색 영역 결정을 위한 기준 및 초기 변이로
 이용하는 깊이 맵 기반 스테레오 정합 기법을 적용하
 였다. 이러한 기법은 가운데 시점에 대한 변이를 추
 정하는데 사용되었으며, 이렇게 함으로써 오정합 확
 률과 정합 과정에 있어서의 계산 시간을 줄일 수 있
 었다. 변이 추정 시에 각 카메라의 파라미터는 Tsai
 의 카메라 보정 기법[7]에 의해 구하였으며, 영상 교
 정(rectification)시에 사용하였다. 또한 구해진 변이 맵
 은 depth 카메라에 의한 깊이 맵과의 융합 과정을 거
 쳐 최종 변이 맵을 생성하게 된다.

3.3 움직임 객체 추출 및 추정

실사 장면의 실제 움직임 객체를 다른 가상 객체로
 대체하거나 정지 실사 배경을 다른 가상 배경으로 대
 치 하는 기능을 제공하기 위해서는 움직임 객체의 위
 치를 정확하게 추적하고, 움직임 객체의 실루엣을 정
 확하게 추출할 수 있어야 한다. 대부분의 기존 연구
 들이 청 스크린을 통한 객체추출 기법을 이용하였으
 나, 본 연구에서는 이웃하는 영상간에 객체의 움직임을
 구하기 위하여 컬러 정보와 깊이 맵내 깊이 변화
 만을 이용하여 객체 움직임 벡터 및 객체 실루엣을
 추출하였다. 이 때 추출된 객체 경계 부분의 보정을
 위하여 Snake 알고리즘을 사용하였다[8]. 추출된 객체
 실루엣 정보를 이용하여 객체 영상에 대하여 depth
 keying 에 의해 합성된 영상을 얻을 수 있다.

3.4 중간 가상 시점 영상 생성

제안한 시스템은 3 시점 카메라에 의해 제한되는
 시점 범위 내에서 앞에서 구해진 변이 정보, 3 시점
 영상, 카메라 정보 등을 이용하여 중간 가상 시점 영
 상 생성을 할 수 있다. 먼저 각 시점에 대한 변이 맵
 을 생성할 가상 시점의 위치로 투영하고, 이 위치에
 3 시점 영상으로부터의 컬러 정보를 투영된 변이 맵
 으로부터 계산된 위치에 채워 넣음으로써 생성될 수
 있다. 생성된 가상 시점 영상은 3 차원 장면 합성 모
 들로 전달되어, 가상 객체를 가상 시점 영상 내 임의
 위치에 삽입하는데 사용할 수 있다..

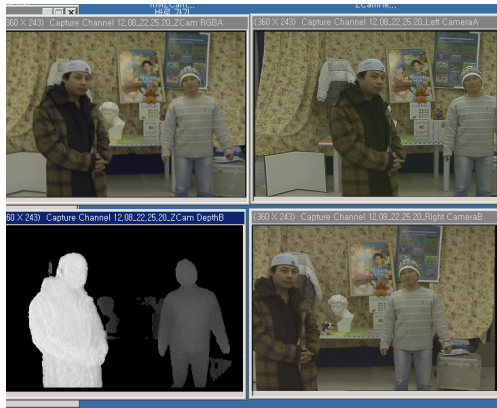
3.5 Depth key 기반 실사/CG 합성

주어진 시점의 영상에 대해 제안한 시스템은 가상
 CG 객체를 특정 깊이에 depth key 에 기반하여 삽입하
 거나, 앞에서 설명한 바와 같이 움직임 객체의 위치
 정보 및 움직임 벡터를 이용하여 배경 대체, 객체 대
 치 등의 콘텐츠 저작 기능도 제공한다. Depth key 합
 성의 장점은 기존 크로마 키에서는 불가능한 배우와
 배경 사이로의 가상 객체 삽입 기능을 제공할 수 있
 다는 것이다.

IV. 시스템 구현 결과

본 절에서는 본 논문에서 제안한 시스템의 구현 결과를

소개한다. 먼저 3 시점 및 depth 카메라 시스템으로부터 3 시점 컬러 영상 및 깊이 맵을 얻었다. 이 때 영상 해상도는 720 X 486 이며, 그림 3(a)에 그 예가 나타나 있다. 그림 3(b)는 가운데 시점에 대한 변이 맵을 보여주고 있으며, 그림 3(c)는 중간시점 영상 생성 결과를, 그림 3(d)는 움직임 객체 추출 결과 및 depth key 를 이용한 실사/CG 합성 결과를 나타내고 있다. 구현 결과, 기본적인 기능은 비교적 잘 구현되었으나, 영상에 따라 변이 정보의 오류에 의해 가상 시점 영상 생성시 일부 화소가 깨지는 경우가 발생하였다.



(a) 3 시점 영상 및 깊이 맵



(b) 가운데 시점 변이 맵



(c) 중간 가상시점 영상 생성 결과



(d) 객체 움직임 추출결과와 Depth key 를 이용한 기반 실사/CG 합성 결과

그림 4. 3 차원 비디오 생성 시스템의 구현 결과

V. 결론 및 추후과제

본 논문에서는 다시점 및 depth 카메라 기반 3 차원 비디오 콘텐츠 생성 시스템을 제안하였으며, 일부 초기 구현 결과를 보였다. 제안한 시스템은 3 시점 영상 및 깊이 맵 획득, 카메라 보정, 다시점 변이 추정, 중간 가상 시점 영상 생성, 움직임 객체 추출/추적을 통한 실사/CG 합성 기능을 제공하는 통합 3 차원 비디오 콘텐츠 생성 시스템이라고 할 수 있다. 추후과제로는 depth 카메라에 의한 깊이 맵을 이용한 정합시점 정밀도를 보다 개선하는 방법 및 가상 시점 영상 생성시 크랙을 줄이는 방법, 렌더링 속도 개선 등이 있다.

참고문헌

- [1] A. Smolić and D. McCutchen, "3DAV Exploration of Video-based Rendering Technology in MPEG," IEEE Trans. on Circuits and Systems for Video Technology," vol. CSVT-14, no.3, pp. 348-356, March, 2004.
- [2] Jong-Il Park, and S. Inoue,"Real-Image-Based Virtual Studio," Proceedings of Virtual Worlds 1998, pp.117-122, 1998.
- [3] O. Grau, "3D sequence generation from multiple cameras," Proceedings of IEEE, International Workshop on Multimedia Signal Processing 2004, Sept., 2004.
- [4] K. Tomiyama, Y. Orihara, M. Katayama, and Y. Iwadate, "Algorithm for Dynamic 3D Object Generation from Multi-viewpoint Images," Proceedings of SPIE, Three-Dimensional TV, Video and Display III, vol. 5599, pp. 153-161, Sept., 2004.
- [5] <http://www.3dvsystems.com>
- [6] Soon-Yong Park and M. Subbarao, "A Range Image Refinement Technique for Multi-view 3D Model Reconstruction," Proceedings of 3DIM'03, pp. 147-154, Banff, Oct., 2003.
- [7] R. Y. Tsai, "Multiframe Image Point Matching and 3-D Surface Reconstruction," IEEE Trans. on Pattern Anal. And Machine Intelligence, vol. PAMI-5, no. 2, pp. 159-173, March, 1983.
- [8] A. Hiton and J. Stark, "Multiple View Reconstruction of People," Proceedings of 3DPVT2004, pp.357-364, Sept., 2004.