

3차원 입체 음향 표출을 위한 다시점 영상에서의 실제 좌표 추정 기술 연구

*정주현 *황동호 **송민기 ***양지희 *박구만

서울과학기술대학교 *전자IT미디어공학과. **일반대학원 미디어IT공학과,
***나노IT디자인융합대학원 정보통신미디어공학전공

*churyel0@naver.com

3D Coordinate Estimation Technical Research in Multi-view Image for 3D Immersive Sound

*Jeong, Ju-Heon *Hwang, Dong-Ho **Song, Min-Ki ***Yang, Ji-Hee *Park, Goo-Man

*Dept. of Electronics and IT Media Engineering,

**Dept. of Media IT Engineering,

***Dept. of Information Technology and Media Engineering,

Seoul National University of Science and Technology

요약

이 논문은 3D 입체 음향에 대한 좌표 취득 및 표출 방법을 제안한다. 3차원 좌표를 추정하기 위해 다시점 영상의 색상 영상을 통해 중간 시점의 깊이 영상을 생성한다. 깊이 영상 내에서 객체에 대한 이미지 좌표와 깊이 정보를 이용하여 3차원 좌표를 취득한다. 실험 결과를 통해 깊이 영상에서 도출된 객체의 좌표를 음원의 좌표로 설정하여 3D 입체 음향 표출의 효율을 높일 수 있음을 확인한다.

1. 서론

최근 자유 선택 시점, 3D 입체 음향과 같은 차세대 실감 미디어 기술이 활발히 개발 중에 있다. 이 중 3D 입체 음향 기술은 대표적인 음향 미들웨어 FMOD가 개발됨에 따라 더욱 편리하게 구현할 수 있게 되었다[1]. 그러나 3D 사운드를 구현하기 위해서는 음원과 청취자의 3D 좌표를 설정해야 한다. 이때 좌표가 실시간으로 불규칙하게 변하는 동적 음원(또는 청취자)의 경우에는 좌표의 설정이 매우 어려워지는 문제가 발생한다. 하지만 3D 입체 음향이 다시점 영상과 결합된 실감 콘텐츠의 경우에는 깊이 영상의 정보들을 입체 음향의 입력 정보로 연계하여 위의 문제점을 해결할 수 있다. 여기서 깊이 영상은 MPEG에서 제공한 DERS(Depth Estimation Reference Software)로부터 생성한다[2,3]. 따라서 DERS로 생성한 다시점 영상의 깊이 정보로부터 음원 또는 청취자의 실제 좌표를 계산으로 도출할 수 있어 FMOD를 통해 정밀한 음향 표현이 가능해진다.

본 논문에서는 두 유형의 실감 콘텐츠사이의 결합으로부터 파라미터 취득과 표출의 효율을 높이는 기술을 제안한다.

2. 본론

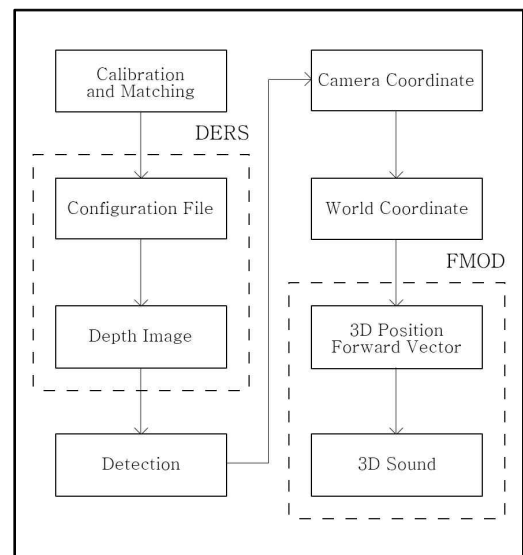


그림 1. 시스템 구성도

그림 1은 본 논문에서 제안하는 깊이 영상에서의 좌표 취득과 3D 입체 음향 표출 기술의 구성도이다.

첫 단계에서 DERS로 다시점 영상의 깊이정보 영상을 생성하기 위해서는 카메라 내부 파라미터와 외부 파라미터를 필요로 한다. 내부 파라미터는 카메라 캘리브레이션을 통하여 초점거리, 주점, 왜곡계수를 구할 수 있으며 외부 파라미터는 2D 영상에서의 픽셀 좌표 $p = (x, y)$ 와 실제 3D상에서의 월드 좌표 $P_w = (X, Y, Z)$ 간의 매칭으로부터 회전 행렬과 이동 벡터를 구할 수 있다[4,5]. 따라서 입력 파라미터와 다시점(왼쪽, 중간, 오른쪽 시점)의 색상 영상 3개로부터 중간 시점의 깊이 영상을 생성할 수 있게 된다.

다음으로 그림 2와 같이 깊이 영상에 영상 이진화를 적용하여 객체와 배경을 분리하고 분리된 객체의 무게중심을 구해 객체의 픽셀 좌표를 검출한다.

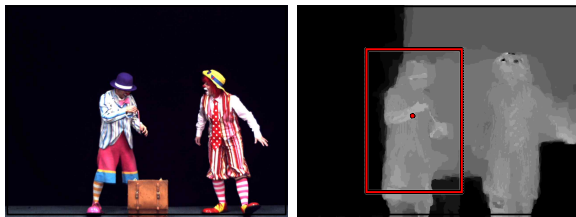


그림 2. 색상 영상과 깊이 영상

그리고 객체의 픽셀 좌표와 이에 대한 깊이 화소 값 v , 깊이 영상에서 도출된 카메라 좌표계 $P_c = (X_c, Y_c, Z_c)$ 상의 특정점인 깊이 값 Z_{near} 값과 Z_{far} 값을 이용하여 식 1로부터 객체의 카메라 좌표 Z_c 값을 도출할 수 있다.

$$Z_c = \frac{1}{\frac{v}{255} \left(\frac{1}{Z_{near}} - \frac{1}{Z_{far}} \right) + \frac{1}{Z_{far}}} \quad \text{식 (1)}$$

따라서 식 2로부터 객체의 카메라 좌표 Z_c 값과 객체의 픽셀 좌표 (x, y) 와 내부 파라미터 A 의 초점거리 f , 주점 c 를 이용해 객체의 카메라 좌표 (X_c, Y_c, Z_c) 를 구할 수 있다.

$$\begin{aligned} x &= f_x \frac{X_c}{Z_c} + c_x \\ \Rightarrow X_c &= \frac{(x - c_x) Z_c}{f_x} \\ y &= f_y \frac{Y_c}{Z_c} + c_y \\ \Rightarrow Y_c &= \frac{(y - c_y) Z_c}{f_y} \end{aligned} \quad \text{식 (2)}$$

마지막으로 식 3을 이용하여 객체의 카메라 좌표를 외부 파라미터 R, T 와 곱하여 객체의 월드 좌표 (X, Y, Z) 를 구할 수 있다[6].

$$\begin{aligned} P_c &= RP_w + T \\ \Rightarrow P_w &= R^{-1}(P_c - T) \end{aligned} \quad \text{식 (3)}$$

다음 단계로 FMOD로 객체의 음향을 표출하기 위해서는 음원과 청취자의 3D 좌표를 필요로 하는데 음원의 3D 좌표는 전 단계에서 취득한 객체의 월드 좌표로 설정한다. 그리고 청취자의 3D 좌표는 카메라의 월드 좌표로 취하며 청취자의 forward 벡터는 카메라 좌표계상의 z축 방향의 기본 벡터를 월드 좌표계상의 벡터로 변환한다[7].

3. 실험결과

본 실험은 Window 10, OpenCV Library 2.4.13, FMOD Library 실험 조건에서 수행되었다. 카메라의 높이는 20cm이며 카메라 간 간격은 4cm이고 카메라와 3차원 기준점 간의 거리는 45cm이다. 또한 3차원 기준점과 객체 간의 거리는 90cm이며 객체의 이동거리는 1m이다.

그림 3과 표 1은 실제 실험한 데이터로부터 DERS를 통해 생성한 한 시점에서의 깊이 영상 내의 객체에 대한 위치 정보를 나타낸 것이다. 객체의 왼쪽, 오른쪽 위치에서의 월드 좌표와 실제 좌표간의 오차를 평균 오차를 공식으로부터 구하여 월드 좌표의 정확성을 평가하였다.

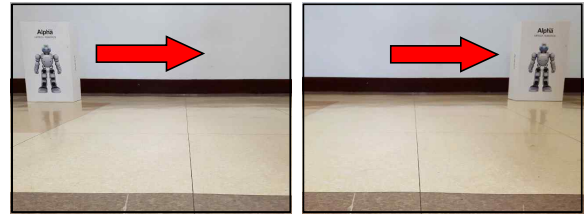


그림 3. 왼쪽, 오른쪽 위치에서의 객체

표 1. 객체의 월드 좌표와 실제 좌표

location	world coordinate (mm)	real coordinate (mm)	error (%)
left	(-200, 34, 954)	(-314, 25, 965)	24.48
right	(660, 29, 948)	(712, 25, 953)	7.94

그림 4는 실험으로부터 얻어낸 객체의 월드 좌표를 FMOD에서 음원의 3D 좌표로 적용시켜 음향을 표출한 것이다. 깊이 영상 내 객체가 카메라에 대해 x축 양의 방향으로 이동함에 따라, 즉 음원이 청취자에 대해 왼쪽에서 오른쪽으로 움직임에 따라 청취자가 듣게 되는 소리의 방향이 변화되는 것을 보여준다.

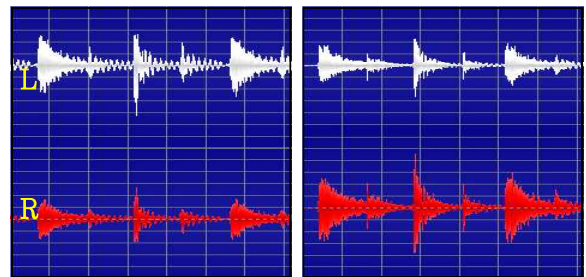


그림 4. 왼쪽, 오른쪽 위치에서의 사운드 스펙트럼

4. 결론 및 향후 연구 방향

현재까지 개발된 입체 음향 표출 기술은 좌표를 수동적으로 취득해야만 하는 문제점을 지니고 있다. 본 논문에서는 3D 입체 영상 및 음향의 효율적인 취득 및 표출을 위해 DERS를 통한 다시점 영상의 깊이 영상으로부터 실시간으로 변하는 음원 또는 청취자의 좌표를 취득하여 FMOD를 통해 입체 음향을 효율적으로 표출할 수 있는 기술을 제안하였다.

향후 연구로 복수 음원들의 정밀하고 안정적인 좌표 취득을 위해 객체 검출 기술을 개선할 예정이다. 그리고 다시점 영상에서 가상 시점을 실시간으로 생성하게 됨에 따라 동시에 청취자의 좌표 또한 실시간으로 취득 및 표출할 수 있도록 할 예정이다.

감사의 글

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2018년도 문화기술 연구개발 지원 사업으로 수행되었음. (R2017030041, 자유선택 시점에서의 문화 콘텐츠 감상 체험 극대화 기술)

참고 문헌

- [1] FMOD Sound System, <http://www.fmod.org> (accessed Sept. 17, 2018).
- [2] M. Tanimoto, T. Fujii and K. Suzuki, "Improvement of Depth Map Estimation and View Synthesis", ISO/IEC JTC1/SC29/WG11, M15090, January 2008.
- [3] M. Tanimoto, T. Fujii and K. Suzuki, "Reference Software of Depth Estimation and View Synthesis for FTV/3DV", ISO/IEC JTC1/SC29/WG11, M15836, October 2008.
- [4] Zhengyou Zhang, "Flexible Camera Calibration by Viewing A Plane from Unknown Orientations", In Proc, International Conference on Computer Vision, September 1999.
- [5] Jean-Yves Bouguet, "Camera Calibration Toolbox for Matlab", 2003.
- [6] G. Golub, C. Van Loan, "Matrix Computation", John Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, Maryland, third edition, 1996.
- [7] R. Hartley, A. Zisserman, "Multiple View Geometry in Computer Vision", Cambridge University Press, 2000.