

새로운 홀 스캔 과 방향성을 고려한 효율적인 영상합성 방법

문지훈, 호요성
광주과학기술원 정보통신공학부
{jhm,hoyo}@gist.ac.kr

초록: 최근 고화질 영상 콘텐츠의 수요 증가와 다시점 컬러 영상과 깊이 영상을 동시에 이용하여 사용자가 원하는 시점에서 시청이 가능하도록 하는 가상 시점 생성 방법에 대한 연구가 활발히 이루어지고 있다. 본 논문에서는 가상 시점 영상 생성시 발생하는 홀 영역을 효율적으로 채우는 방법을 제안한다. 컬러 영상과 깊이 영상의 위치가 정확히 맞지 않기 때문에 합성된 영상에서는 홀 영역이 나타난다. 홀은 기준 시점 영상에서 보상할 수 없는 가려짐 영역(occlusion)으로 정의된다. 제안하는 방법을 수행하기 위해 우선 경계 잡음 영역을 제거한다. 경계 잡음 영역을 홀 영역으로 인식해 처리할 수 있도록 홀 영역을 확장하여 경계 잡음을 제거한다. 홀 영역을 채우기 위해 홀 주변 영역을 지그재그 스캔 방법(Zigzag Scan method)을 이용하여 채워나가게 된다. 실험을 통해 제안한 홀 채움 방법이 기존의 홀 채움 방법보다 좋은 성능을 나타냄을 확인할 수 있었다.

주제어: 3D-Warping, Image Synthesis, Hole filling, Free-view point TV, 3차원 워핑, 영상합성, 홀 채움, 자유시점 TV

I. 서론

3차원 비디오(3D video) 시스템은 다시점 비디오를 기반으로 하여 시청자에게 보다 실감나는 콘텐츠를 제공하는 시스템을 말한다. 이로 인해 점차 3차원 비디오 시스템에 대한 수요가 늘어가고 있는 추세이다. 3D 산업은 영화, 게임, 방송 등 다양한 멀티미디어 적용 분야에서 새로운 미디어 서비스로 자리잡아가고 있다. 이 콘텐츠는 오토 스테레오스코픽(autostereoscopic) 장치를 이용하거나 자유시점 TV에 이용될 수 있다. 이를 통하여 사용자는 시청 시점을 자유롭게 선택하여 볼 수 있게 됨으로써 보다 실감나는 3차원 입체영상을 감상할 수 있다. 이러한 3차원 영상 산업을 위한 시스템은 이미 많은 연구를 통해 기술의 미래성과 창출효과에 대해 입증되어 있다. 유럽에서는 ATTEST (Advanced Three dimensional Television System Technologies)를 통해 2002년부터 연구를 지속해오고 있으며, 미국의 NASA와 MIT에서는 3D 영상처리 기술과 홀로그래피(holography) 기술이 각각 개발되고 있는 중이다. 일본의 나고야(Nagoya) 대학에서는 광성 공간(Ray-Space)를 이용한 자유시점 TV를 연구하고 있는 중이다. 하지

만 최근에는 3차원 영상을 시청하기 위해 착용해야 하는 안경의 불편함, 시각적 피로감, 3차원 영상을 지원하는 콘텐츠의 부재 등으로 기대보다는 3차원 영상 산업이 활성화 되지 않고 있다는 평이 나오고 있는 상황이다. 이는 여러 연구 성과들이 하나의 시스템으로 융합되지 못하고 이를 쉽게 적용할 수 있는 방법들의 부재 때문에 나타나는 현상이라고 볼 수 있다. 3차원 영상 산업의 활성화를 위해서는 2차원 콘텐츠와 같이 수가 다양하고 손쉽게 획득할 수 있어야 한다. 그리고 안경과 같은 보조 장비 없이 편안하게 3차원 영상을 즐길 수 있어야 하며 눈의 피로를 최소화 해야 한다.^[1]

일반적으로 스테레오로 이루어진 3차원 영상을 획득하기 위해 두 대의 카메라를 사용한다. 스테레오 카메라로 촬영된 영상을 통해 입체감을 느끼기 위해서는 안경을 통해 양안시차를 조절할 수 있고, 또한 안경 없이 다시점 영상을 제공하여 어느 시점에서든 시청자가 입체감을 느낄 수 있게 하는 방법이 있다. 최근 MPEG에서 새로운 3차원 부호화 기술의 개발 연구에 따라 다시점 비디오 기술과 자유시점 제공 기술에 대한 연구가 진행되었다. 여러 대의 카메라를 통해 전달되는 정보의 전송 효율을 높이기 위해 먼저 수신단에서 다시점 색상 영상과 각 색상 영상에 대응하는 깊이 데이터를 획득하고, 이를 부호화해 전송한다. 전송 받은 깊이 데이터를 기반으로 하여 중간 영상 생성을 통해 3차원 비디오 콘텐츠를 제작하게 된다. 이를 위해 MPEG에서는 가상 시점 생성 소프트웨어(VSRS: view synthesis reference software)와 깊이 추정 소프트웨어(DERS: depth estimation reference software)를 이용하고 있다.^[2]

II. 3차원 비디오 시스템

고화질의 가상 시점 영상합성 방법이 3차원 비디오 시스템에서 중요한 이유를 설명하기 위해 먼저 이 시스템의 정의와 구조에 대해 이해 해야 한다. 이 장에서는 현재 MPEG에서 표준화를 진행하고 있는 3차원 비디오 시스템의 정의와 구조에 대해 간략하게 설명 한다.

그림 1에 나타난 것처럼 3차원 비디오 시스템은 다시점 컬러 영상들의 입력과 이에 해당하는 깊이 영상을 입력 받게 된다. 입력 받은 영상들은 부호화를 진행해 전송되고 이 부호화된 정보를 수신하는 기기들은 비트 스트림을 복호화 하여 본래의 콘텐츠로 복원하는 방법을 사용한다. 복원된 영상은 본 논문에서 중점적으로 다룰 중간시점 영상합성 방법을 이용하여 중간시점 영상을 생성한다.

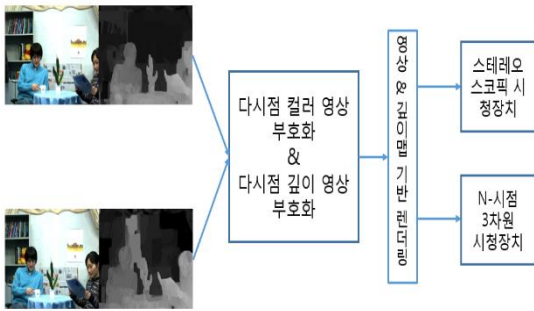


그림 1. 3차원 비디오 재생 시스템

이때 시청자가 사용하는 시청 기기에 따라 중간시점의 생성 개수가 달라진다. 스테레오 스코픽 시청장치를 이용할 경우에는 깊이감을 가장 효율적으로 느낄 수 있는 최적의 두 시점의 영상을 합성하여 가상 시점 영상을 생성하여 시청자에게 보여준다. 이때 생성되는 가상시점 영상은 VSRS와 같은 가상시점 영상합성 방법을 사용하여 생성하게 된다. 3차원 비디오 시스템은 제한된 수의 다시점 영상을 사용하여 보다 많은 수의 다양한 시점을 생성하고 시청 장치에 따라 최적의 영상을 선택하여 보여주는 시스템이라고 할 수 있다.

III. 3D-warping 을 이용한 가상시점 영상합성 방법

1. 3차원 워핑(warping)

이번 장에서는 2장에서 설명한 가상시점영상의 생성 방법에 대해 전체적으로 설명하고, 가상시점 영상합성에 있어서 중요한 3D-warping방법에 대해 자세히 살펴본다.

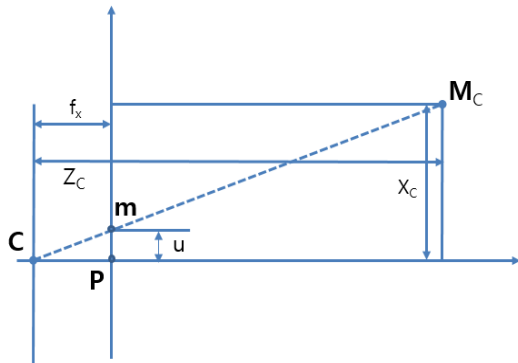


그림 2. 카메라 영상의 투영

그림2는 핀홀 카메라 모델의 영상 투영 구조를 2D 형태로 표현한 것이다. 핀홀 카메라로 촬영한 영상은 주로 x축의 f_x 에 역으로 상이 맺히게 된다. 카메라를 통해 촬영한 영상은 2D구조로 저장되지만 실제로 촬영된 사물들은 3D구조상에 위치해 있으므로 정확한 위치를 정의하기 어렵다. 이 때문에 영상이 맺히는 위치를 2D 구조의 f_x 의 위치로 옮겨 해석한다. f_x 의 위치를 카메라 초점거리(focal length)라고 한다. 실제 카메라 좌표계(real camera coordinate)에서 물체의 3D 좌표가 영상 평면(image plane)에 투영되어 상이 맺히는 위치는 삼각 비례법을 사용하여 알아낼 수 있다.

각 비례법을 사용하여 알아낼 수 있다.

$$\frac{u}{X_c} = \frac{f_x}{Z_c} \quad (1)$$

그림 2에 나타나 있는 변수들을 삼각 비례법으로 정리하면 식 (1)과 같이 나타낼 수 있다. 이를 u 에 대해 식 (2)와 같이 정리하면 실제 카메라 영상이 맺히는 위치를 파악할 수 있다.

$$u = \frac{f_x X_c}{Z_c} \quad (2)$$

카메라 좌표계C의 시점에서 보면 초점거리 P의 위치가 $(0,0)^T$ 가 아니므로 원점의 위치를 찾아내기 위해 보정할 수 있도록 보정 변수를 사용하여 식 (3)과 같이 나타낼 수 있다.

$$u - p_x = \frac{f_x X_c}{Z_c} \\ \therefore u = \frac{f_x X_c}{Z_c} + p_x, v = \frac{f_y Y_c}{Z_c} + p_y \quad (3)$$

3차원 워핑은 지금까지 설명한 원리를 기반으로 하여 수행된다. 임의의 가상 시점 영상을 생성하기 위해서는 참조영상을 목표 시점인 가상시점으로 시점을 이동시켜야 한다. 즉, 참조시점에 있는 객체를 가상시점 영상으로 옮겨야 하는 과정을 거쳐야 하는 것이다. MacMillan이 제안한 3차원 워핑을 이용한 중간시점 영상의 생성 방법은 카메라 파라미터와 깊이 정보를 이용하여 참조시점의 각 화소가 가상 시점영상의 어느 위치에 생성되는지 정의하여 중간시점 영상을 생성하게 된다.^[3]

우선 3차원 실세계 좌표계를 카메라 좌표계로 이동하기 위해서는 회전(rotation)과 이동(translation) 변환이 필요하다. 이 두 가지 변환은 카메라의 외부 파라미터로 정의된다. 카메라 파라미터는 카메라 좌표계와 세계 좌표계 간의 관계를 기술한 값으로, 카메라 내부 파라미터와 외부 파라미터 이렇게 두 가지로 정의된다. 내부 파라미터 값은 그림 2에서 구한 실제 촬영된 영상이 맺히는 상의 위치 f_x 를 포함하는 정보로 정의된다. 지금까지 설명한 내용을 식 (4)와 같이 표현 할 수 있다.

$$Z_c \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} f_x & 0 & p_x \\ 0 & f_y & p_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} = A \cdot M_c \quad (4)$$

식 (4)에서 A는 카메라 내부 파라미터를 나타내고 M_c 는 카메라 외부 파라미터를 나타낸다. M_c 를 회전 변환 R과, 이동 변환을 t로 그리고 $Z_c(u,v,1)^T$ 를 m' 으로 정리하면 식 (5)와 같이 표현할 수 있다.

$$m' = A \cdot R \cdot M_w + A \cdot t \quad (5)$$

식 (5)를 이용하여 가상시점 영상의 한 화소를 깊이 값 d_r 을 이용하여 세계 좌표계로 변환할 수 있게 된다. 이 관계는 식 (6)과 같이 나타낼 수 있다.

$$M = R^{-1} \cdot A^{-1} \cdot m_s \cdot d_r - R^{-1} \cdot t \quad (6)$$

M은 깊이 정보를 이용하여 카메라로 촬영된 영상을 실제 좌표계(world coordinate)로 변환된 결과를 의미한다. 실제 좌표계에서 다시 영상 평면에 투영하는 과정이 있지만 가상시점 영상합성은 M을 구하는 과정까지만 진행하는 것으로 가능하다. 3차원 워핑의 전체적인 구조는 그림 3에 나타나 있다.

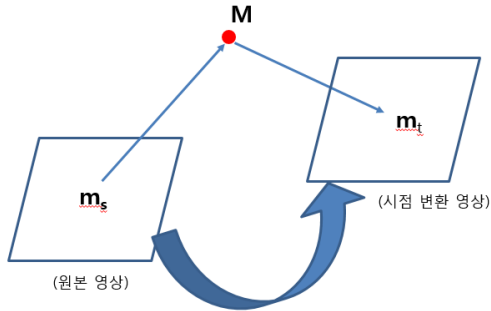


그림 3. 3차원 워핑의 구조

그림 4는 좌영상과 우영상을 가상시점으로 이동한 결과를 나타낸 것이다. 결과에 보이는 것 같이 각 영상들을 가상시점으로 이동하게 되면 완전한 영상이 생성되지 않고 빈 부분이 발생하는 것을 확인할 수 있다. 이를 3차원 워핑 결과 발생하는 홀(hole)이라고 하며 홀을 효율적으로 채우는 것이 가상 시점 영상을 생성하는데 중요한 요소이다.



그림 4. 참조 영상의 가상시점 이동 결과

2. 홀 채움(Hole Filling) 방법

3차원 워핑을 이용하여 시점을 이동하면 좌영상과 우영상에서 보이지 않던 영역이 가상시점에 새롭게 드러나게 되면서 이 부분은 채워지지 않은 빈 영역으로 나타나게 된다. 이러한 영역을 비폐색영역(disocclusion) 또는 홀(hole) 영역이라고 한다. 이러한 영역은 참조영상에는 없는 정보이므로 이를 채우는 것이 중요한 문제이다. 가상 시점영상은 다시점 영상을 기반으로 만들기 때문에 3차원 워핑 결과 발생하는 홀 영역을 다른 참조영상에서 가져와 채울 수 있다. 기존의 인페인팅 방법^[4]를 이용한 홀 채움 결과를 그림 5에서 확인할 수 있다.



그림 5. 기존 홀 채움 결과

IV. 제안하는 홀 채움 방법

기존 VSRS에서 사용하는 홀 채움 방법은 홀 발생 영역 주변의 가져다 사용하는 방법으로 구성되어 있다. 홀 영역은 주로 객체의 경계 부분에서 발생하게 된다. 하지만 홀 주변 영역 화소값을 사용하게 된다면 객체의 정보와 배경의 정보가 홀을 채우는데 사용되게 된다.

이러한 방법으로 홀을 채우게 되면 채워진 홀 영역에서 이미지가 흐릿하게 보이는 현상(blur effect)가 발생한다. 본 논문에서는 이러한 잡음 현상을 제거하기 위해 지그재그 가중치 평균 방법을 제안한다. 그 전에 홀 주변 영역에서 어느 부분이 배경인지 또는 객체인지 구분해야 한다. 이는 깊이 영상을 이용한 정보를 사용하면 쉽게 알 수 있다. 그 예시를 그림 6에서 보이고 있다.

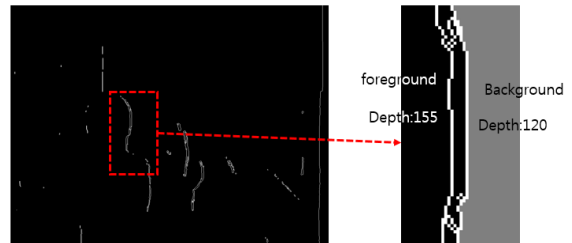


그림 6. 배경과 전경의 분류

배경의 깊이 값이 120이고 전경의 깊이 값이 155라면, 우리는 155의 깊이 값을 갖는 부분이 전경임을 판단할 수 있다. 홀 채움 과정은 배경이라고 판단되는 부분에서부터 지그재그 스캔을 하며 홀을 채울 값을 결정하게 된다. 홀 영역을 기준으로 좌측과 우측 각각 배경일 경우를 고려하여 좌 우측을 고려하여 스캔 방향을 설정해준다. 이 과정을 그림 7에서 보이고 있다.

지그재그 스캔만을 통해 홀을 채울 값을 구하게 된다면 평균값을 구하는 것과 같기 때문에 홀을 채우게 되면 번짐 현상이 나타나게 된다.

이를 방지하기 위해 홀 영역의 방향성도 고려해준다. 홀을 중심으로 하여 4개의 방향 후보를 가지고 최적의 화소 값을 찾을 수 있는 비용 함수를 생성한다. 최적의 값은 각 방향에 따라 홀 주변의 색상 값의 차이가 가장 작은 값을 의미한다. 4개의 방향 후보들은 그림 8에 나타나 있다.

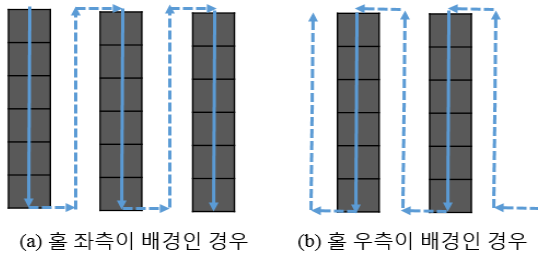


그림 7 홀 영역 스캔 방향

홀 영역 스캔과 마찬가지로 배경영역이 좌측과 우측에 있을 경우를 고려하여 두 가지 방향 후보그룹을 생성한다.

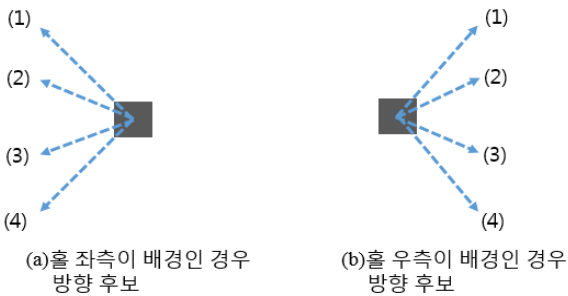


그림 8. 방향성 고려 후보 그룹

방향성과 홀 영역 스캔방향을 고려한 비용 함수는 식 (7)과 같이 나타낼 수 있다.

$$Cost = \frac{\sum_{x,y \in Neighbor} |ZS(x,y) - DC(x,y)|}{NeighborNumber} \quad (7)$$

ZS는 지그재그 스캔을 수행하게 되는 홀 주변 영역을 의미하며, DC는 홀 이웃 영역의 후보군의 값들을 나타낸다. 식 (7)의 비용 값을 가장 적게 갖게 하는 방향이 최종적으로 홀 영역을 채우게 될 화소 값으로 결정된다. 이 과정은 식 (8)에 나타나 있다.

$$Direction = \arg \min \{ Cost \} \quad (8)$$

V. 실험 결과 및 분석

지금까지 제안한 방법들의 성능을 평가하기 위해 MPEG에서 제공하는 다시점 영상 시퀀스인 'Mobile'과 'NewsPper'를 실험 영상으로 사용한다. 전경과 배경을 구분하기 위해 사용하는 문턱치 값은 7을 사용했다.

그림 9와 그림 10의 실험 결과는 지그재그 스캔만을 이용하여 홀 영역을 채운 결과와, 방향성을 고려하여 비용함수를 사용해 홀 영역을 채운 결과를 보여주고 있다. 지그재그 스캔을 통해 얻은 결과는 그림 9와 그림 10의 중간 영상이고 방향성까지 고려한 결과는 맨 오른쪽에 있는 영상이다. 방향성을 고려한 홀 채움 결과 채워진 영역에 흐릿한 현상이 나타나지 않고 배경 부분과 어색하지 않게 결과가 나타남을 확인할 수 있다.



그림 9. 'Mobile'영상의 기존, 제안한 방법을 통해 복원된 영상

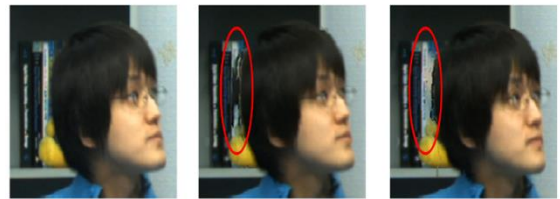


그림 10. 'NewsPaper'영상의 기존, 제안한 방법을 통해 복원된 영상

VI. 결론

본 논문에서는 3차원 비디오의 가상시점 영상합성 중 3차원 워핑 결과 발생하는 홀 영역을 효율적으로 채우는 방법에 대해 제안한다. 지그재그 스캔 방법과 방향성을 고려한 홀 채움 방법을 사용한다. 지그재그 스캔만을 이용한 경우에는 약간의 흐림 현상이 발생하지만 방향성을 동시에 고려해준 경우에는 그 성능이 더 개선됨을 확인할 수 있었다. 현재 다시점 비디오 부호화의 테스트 영상과 3차원 비디오 테스트 영상을 이용하여 실험한 결과, 제안한 방법이 기존의 홀 채움 방법보다 더 나은 성능을 나타냄을 확인했다.

감사의 글

이 논문은 2013년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2013-067321)

참고문헌

[1] G.M. Um, G.H. Cheong, W.S. Cheong, N.H. Hur, "Technical Development and Standardization Trends of Multi-view 3D and Free-view point Vide," *The Magazine of the IEEK*, vol.38, no. 2, pp. 18-23, Feb., 2011.

[2] ISO/IEC JTC1/SC29/WG11, "Introduction to 3D Video," June, 2009.

[3] L. McMillan, "An image-based approach to three-dimensional computer graphics Technical report," *Ph.D. Dissertation, UNC Computer Science*, 1999.

[4] A. Telea, "An image in-painting technique based on the fast marching method," *J. Graphics Tools*, vol.9, no. 1, pp. 25-36, 2004.