

정면시점 영상합성 방법을 통한 시선 맞춤 방법

이상범¹, 신인용¹, 양승준², 이한규², 차지훈², 호요성¹

광주과학기술원¹, 한국전자통신연구원²

sblee@gist.ac.kr, siy0808@gist.ac.kr, sjyang@etri.re.kr, hkl@etri.re.kr, jihun@etri.re.kr, hoyo@gist.ac.kr

요약

본 논문에서는 정면시점 영상합성 방법을 통한 시선 맞춤 (eye contact) 방법을 제안한다. 제안하는 방법은 수렴형 배열로 구성된 스테레오 카메라를 이용해서 화자를 촬영한 후, 카메라 매개변수 획득 및 카메라 교정 등의 영상 전처리 과정을 거친다. 전처리 과정이 끝난 스테레오 영상은 깊이 탐색 방법에 이용되고 이 방법을 통해 카메라와 화자 사이의 거리를 계산할 수 있다. 최종적으로 정면시점 영상합성 방법을 이용해서 화자가 정면을 바라보는 듯한 영상을 획득할 수 있다. 컴퓨터 모의 실험 결과를 통해, 제안하는 방법이 자연스러운 정면시점 영상을 합성할 수 있었고, 원격의 화자와 시선 맞춤이 가능한 것을 확인했다.

1. 서론

“3D가 디지털 미디어의 르네상스를 만들어냈다. 3D는 엔터테인먼트의 획기적 사건이자 전 세계적 변화이다.” 영화 “아바타”로 3D의 전 세계적인 관심을 이끌어 낸 제임스 카메론 감독이 제7회 서울 디지털 포럼에서 언급한 내용이다 [1]. 최근 폭발적인 증가세를 보이고 있는 3D 시장의 점화제 역할을 한 제임스 카메론 감독의 발언은 방송 시스템이 아날로그에서 디지털로 전환되면서 영상 산업 전반에 큰 변화를 가져 왔듯이, 가까운 미래에 디지털 미디어가 2차원에서 3차원으로 전환되는 영상 산업의 또 다른 혁명을 불러일으킬 것이라는 전망과 일치한다 [2]. 실제로, 세계 선진국에서는 3차원 방송을 위한 3차원 영상 콘텐츠를 제작하고 있으며, 국내에서도 여러 방송 사업자들을 중심으로 3D 시험방송을 준비하고 있다 [3].

최근 MPEG (moving picture experts group) 국제 표준화 기구에서는 3차원 비디오 시스템을 정의하고, 다시점 색상영상과 깊이영상을 포함하는 3차원 비디오를 압축 부호화하는 국제 표준화 작업을 진행하고 있다 [4]. MPEG에서 정의한 3차원 비디오 시스템은 3시점 혹은 그 이상의 넓은 시야각을 제공하는 고해상도의 3차원 비디오 시스템을 말한다. 3차원 비디오 시스템을 구현하기 위해서는 다수의 카메라로 획득한 넓은 시야각의 다시점 영상을 이용해서 3차원 장면의 거리 정보를 표현하는 깊이영상을 추정하는 기술과 깊이영상을 이용하여 사용자가 원하는 임의의 시점에서 장면을 시청할 수 있도록 중간시점 영상합성 기술이 사용된다. 그림 1은 MPEG에서 구상하는 3차원 비디오 시스템의 개념도를 나타낸다.

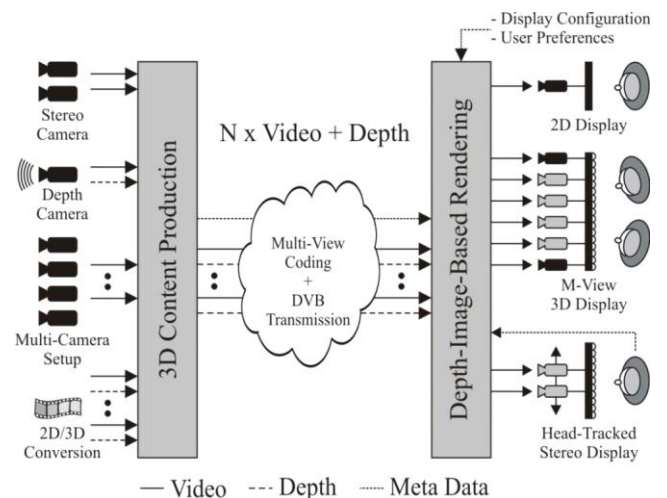


그림 1. 3차원 비디오 시스템

이러한 3차원 비디오 시스템의 핵심 기술들 가운데 하나인 깊이 탐색 기술과 영상 합성 기술은 다양한 응용 분야에 이용될 수 있는데, 그 대표적인 예가 원격 화상 회의를 위한 시선 맞춤 (eye contact) 기술이다. 최근, 독일의 HHI 연구소에서는 앞서 언급한 주요 기술들을 이용한 3차원 원격 화상회의 시스템을 개발했다 [5]. 이 방법은 4대의 카메라를 이용해서 화자의 깊이 정보를 탐색한 다음, 영상 합성 과정을 통해 원격의 화자의 시선을 맞춘다. 하지만, 이 방법은 그 성능에 비해 하드웨어 구성이 너무 복잡하고 시스템을 구축하는데 너무 많은 비용이 든다는 단점이 있다.

본 논문에서는 정면시점 영상합성 방법을 통한 시점 조정 방법을 제안한다. 이 방법은 기존의 방법과는 달리 수렴형 배열로 구성된 두 대의 스테레오 카메라를 이용해서 정면시점 영상을

합성한다. 일단 화자를 촬영한 후, 카메라 매개변수 획득 및 카메라 교정 등의 영상 전처리 과정을 거친다. 전처리 과정이 끝난 영상은 깊이 탐색 방법에 이용되고 이 방법을 통해 카메라와 화자와의 거리를 계산할 수 있다. 최종적으로 정면시점 영상합성 방법을 이용해서 화자가 정면을 바라보는 듯한 영상을 획득할 수 있게 된다.

2. 3차원 비디오 시스템을 이용한 정면시점 영상합성 기술

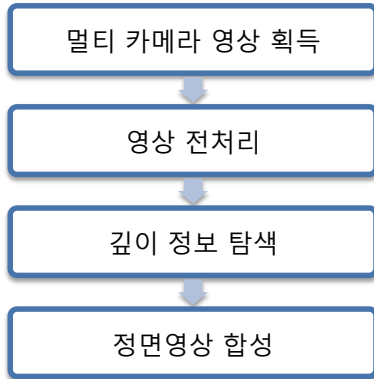


그림 2. 제안하는 방법의 전체 블록 다이어그램

그림 2는 본 논문에서 제안하는 방법의 전체 블록 다이어그램을 나타낸다. 수렴형 배열로 구성된 멀티 카메라를 이용해서 화자를 촬영한 후, 영상 전처리 과정을 거친다. 그 다음, 깊이 탐색 방법을 통해 카메라와 화자와의 거리를 계산한다. 끝으로 정면시점 영상합성 방법을 이용해서 화자의 시점을 변경한다.

2.1 영상 획득 및 전처리

카메라 보정(camera calibration)은 카메라 매개변수를 예측하는 기술이다. 격자 무늬의 패턴이 촬영된 여러 장의 2차원 영상들에서 추출한 특징점을 기반으로 카메라의 내부 매개변수와 외부 매개변수를 계산한다. 카메라 내부 매개변수는 카메라의 초점 거리와 같이 카메라 내부의 물리적 특성을 나타내는 값들로 이루어진 행렬로 표현되며, 외부 매개변수는 3차원 공간에서 카메라의 방향과 위치를 나타내는 회전 행렬과 이동 벡터로 이루어진다. 카메라의 내부와 외부 매개변수를 이용하여 카메라의 투영 행렬을 구할 수 있고, 이 투영 행렬은 3차원 공간상의 한 점을 2차원 영상 평면상의 한 점으로 옮겨오는 역할을 한다.

카메라 보정을 통해 얻은 카메라 변수 혹은 카메라 투영 행렬은 3차원 영상 처리 및 응용에서 가장 기본이 되는 필수적인 정보이다 [6]. 그림 3은 카메라 보정을 위해 좌시점에서 촬영한 패턴 영상을 나타낸다. 여러 대의 카메라를 사용하는 경우, 모든 카메라에 대해 보정을 수행한다.

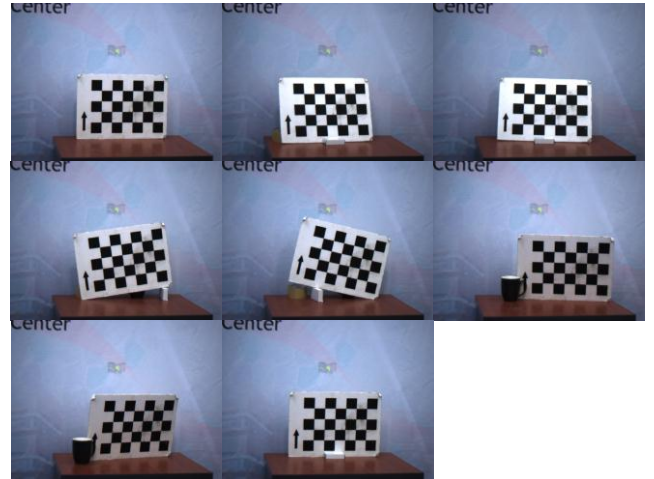


그림 3. 카메라 보정을 위해 촬영한 패턴 영상

인접한 여러 대의 카메라로 촬영된 영상에는 일반적으로 기하학적 오차가 존재한다. 이 오차는 카메라의 배열을 수동으로 구성한 것에서 기인한 것으로, 각 시점의 영상에서의 대응점들의 수직 좌표와 수평 방향으로의 시차인 변이가 일정하지 않은 것으로 나타난다. 또한 같은 카메라를 사용할 때에도 카메라 보정을 통해서 얻어지는 각각의 카메라 내부 변수들 사이에 역시 오차가 존재하게 된다. 이러한 오차들은 깊이 영상을 생성하고 중간 시점 영상을 합성함에 있어서 품질을 떨어뜨리는 원인이 된다.

다시점 영상 정렬(image rectification)은 카메라 매개변수를 이용하여 구한 변환식을 각 시점의 영상에 적용하여 기하학적 오차를 최소화하는 작업이다. 카메라 매개변수로부터 카메라의 광축을 예측할 수 있으며, 정렬되지 않은 광축은 영상 정렬 방법을 통해 정렬될 수 있다. 정렬된 다시점 영상은 대응점 간에 수직 방향으로의 불일치 없이 수평 방향으로의 변이만 가지게 된다. 그림 4는 각각 정렬되기 전의 영상과 정렬된 후의 영상을 나타낸다 [7].



(a) 영상 정렬 전



(b) 영상 정렬 후

그림 4. 영상 정렬 결과

2.2 깊이 정보 탐색 방법

깊이 영상(depth map)이란 영상 내에 존재하는 객체들의 3차원 거리 정보를 8비트로 표현한 영상을 말하며 깊이 영상의 화소값은 각 해당 화소의 깊이 정보를 나타낸다. 깊이 영상은 깊이 카메라를 이용해 직접 획득할 수 있지만 스테레오 카메라 및 다시점 카메라를 이용하는 경우에는 깊이 영상을 계산적으로 추정해서 얻는다. 다시점 깊이 영상을 획득하기 위해서는 다시점 영상 각 시점의 상관관계를 이용하여 깊이 정보를 계산적으로 탐색하는 스테레오 정합 기술을 가장 많이 이용한다. 스테레오 정합 기술은 인접한 두 시점 영상에서 객체가 수평으로 이동한 정도, 즉, 변이(disparity)를 계산해서 깊이 정보를 획득하는 방법이다 [8]. 이 방법은 특별한 센서 없이도 깊이 정보를 획득할 수 있기 때문에 비용이 적게 들고, 이미 촬영된 영상에 대해서도 깊이 정보를 획득할 수 있다는 장점을 지닌다.

변이값을 계산하기 위해서는 기준이 되는 좌영상의 모든 화소에 대해 우영상에서 동일한 위치의 화소를 탐색해야 하는데 이를 위해서 정합 함수를 사용한다. 정합 함수는 두 시점의 화소를 비교했을 때의 오차값을 나타내는 것으로 오차값이 작을수록 동일한 위치일 확률이 높아진다. 깊이 탐색을 위한 정합 함수는 다음과 같이 정의된다.

$$E(x, y, d) = E_{data}(x, y, d) + \lambda E_{smooth}(x, y, d) \quad (1)$$

$$E_{data}(x, y, d) = |I_L(x, y) - I_R(x - d, y)| \quad (2)$$

$$E_{smooth}(x, y, d) = \sum_{(x_i, y_i) \in N_p} |D(x, y, d) - D(x_i, y_i, d)| \quad (3)$$

여기서 (x,y)는 비교를 위한 영상의 화소의 좌표를 나타내고, d는 탐색 범위 내의 구하고자 하는 깊이값을 나타낸다. E_{data}(x,y,d)는 좌영상과 우영상의 화소값의 차이를 나타낸다. E_{smooth}(x,y,d)는 깊이 영상 내의 인접 화소들과의 깊이값의 차이를 나타낸다. 위와 같은 정합 함수를 이용해서 좌우 시점 각각에 대한 깊이 영상을 탐색한 결과는 그림 5와 같다.

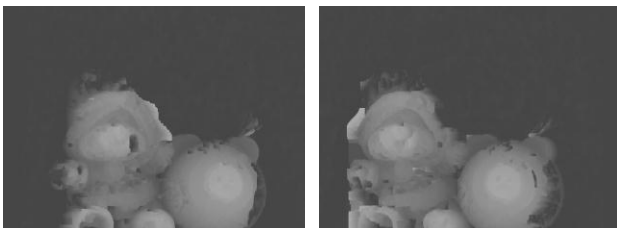


그림 5. 깊이 정보 탐색 결과

2.3 정면시점 영상합성 방법

정면시점 영상합성은 그림 6과 같이 크게 3단계로 나눌 수 있다. 첫 번째는 시점 이동

과정이다. 시점 이동이란 앞에서 구한 깊이 정보를 이용하여 색상영상을 두 시점의 가운데 위치하는 가상의 시점으로 투영하는 방법을 말한다. 시점 이동으로 인해 참조 시점에서 존재하지 않았던 영역은 홀(hole)로 나타나게 된다. 홀은 좌, 우 참조 화면으로부터 중간시점으로 이동된 두 영상을 하나로 합치는 영상 통합 과정을 통해 대부분 채워진다. 마지막 과정은 영상 통합 과정에서 남아 있는 홀을 영상 보간법이나 인페인팅 (inpainting) 기술을 통해 채우는 과정이다.

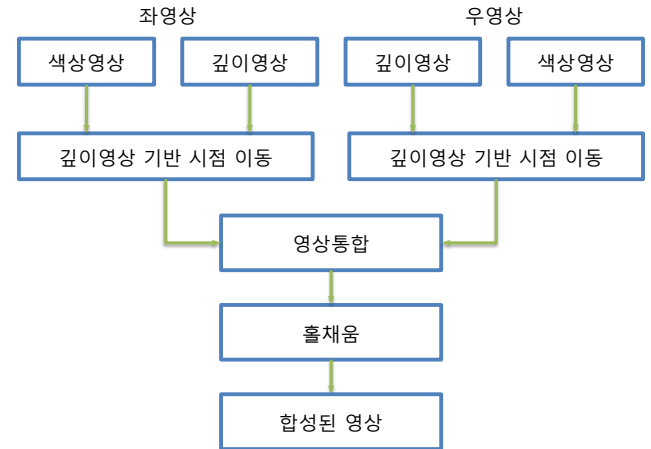


그림 6. 정면시점 영상합성 방법

그림 5에서 얻은 깊이영상을 이용해서 정면시점 영상을 합성한 결과는 그림7과 같다.



그림 7. 정면시점 영상합성 결과

3. 실험 결과 및 분석

본 논문에서 제안하는 시선 맞춤 방법의 성능을 평가하기 위해, 그림 8과 같은 시스템을 구축하고 영상을 획득했다. 카메라의 간격은 42인치 디스플레이를 기준으로 했을 때, 약 93cm, 시정거리는 2.7m로 설정했다. 영상을 획득하기 위해서 배경의 변화는 없고 배경과 화자의 거리는 거의 차이가 없도록 제한사항을 두었다. 카메라

모델은 Point Grey Research 사의 Grasshopper이며 해상도는 1280x960으로 설정했다.



그림 8. 시선 맞춤을 위한 스테레오 카메라 시스템

그림 9는 앞에서 언급한 주요 기술들을 이용해서 얻은 결과물을 나타낸다. 각 단계별로 얻은 영상들을 이용해서 최종적으로 정면시점 영상을 합성하는데, 그림 9에서 알 수 있듯이, 제안하는 방법을 통해 합성한 정면시점 영상이 자연스러운 것을 확인할 수 있다.



(a) 색상영상



(b) 깊이영상



(c) 정면시점 영상

그림 9. 영상합성 결과

4. 결론

본 논문에서는 정면시점 영상합성 방법을 통한 시선 맞춤 방법을 제안했다. 제안하는 방법은 기존의 방법과는 달리 수렴형 배열로 구성된 두대의 카메라만을 이용해서 정면시점 영상을 합성했기 때문에 기존의 방법에 비해 하드웨어 구성의 복잡도가 단순하면서 시선 맞춤 문제를 해결할 수 있었다. 또한, 3차원 비디오 시스템의 주요 기술인 깊이 정보 탐색 및 정면시점 영상합성 방법을 이용해서 자연스러운 시선 맞춤이 가능한 것을 확인할 수 있었다.

감사의 글

본 연구는 지식경제부 및 방송통신위원회의 IT 산업 원천기술개발사업의 일환으로 수행하였음. [KI002070, IPTV용 Interactive 시점제어 기술 개발]

참고문헌

- [1] “제임스캐머런 ‘상상력, 기술의 新 르네상스,’” <http://www.mt.co.kr/view/mtview.php?type=1&no=2010051308483641428&outlink=1>.
- [2] 호요성, 오관정, “다시점 비디오 기반의 3차원 실감방송,” 방송과 기술, 제 161권, pp. 136-143, 2010. 05.
- [3] “방통위, 3차원방송 등 집중 육성키로,” http://www.hani.co.kr/arti/economy/economy_general/419977.html.
- [4] 호요성, 이상범, “3차원 비디오 부호화 기술의 국제 표준화 동향,” TTA Journal, No. 123, pp. 77-83, 2009. 06.
- [5] P. Kauff, O. Schreer, “An immersive 3D video-conferencing system using shared virtual team user environments,” in Proc. of international conference on Collaborative virtual environments, pp. 105-112, Sep. 2002.
- [6] Z. Zhang, “A Flexible New Technique for Camera Calibration,” IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 22, no. 11, pp. 1330-1334, 2000.
- [7] 강윤석, 호요성, “평행 카메라 배열에서 촬영된 다시점 영상의 효율적인 정렬화 방법,” 제21회 신호처리합동학술대회 논문집, 제21권, pp. 92-95, 2008.
- [8] D. Sharstein and R. Szeliski, “A Taxonomy and Evaluation of Dense Two-frame Stereo Correspondence Algorithms,” Proc. of IEEE Workshop on Stereo and Multi-Baseline Vision, pp. 131-140, April 2001.