깊이 영상의 후처리 필터링 기술을 통한 원격 화상 회의의 시선 맞춤 방법

이상범, 양승준*, 호요성
광주과학기술원, *한국전자통신연구원

sblee@gist.ac.kr, *sjyang@etri.re.kr, hoyo@gist.ac.kr

Eye Gaze Correction for Teleconferencing Using Depth Video Filtering

Sang-Beom Lee, Seung-Jun Yang*, and Yo-Sung Ho
Gwangju Institute of Science and Technology (GIST),
*Electronics and Telecommunications Research Institute (ETRI)

요 약

본 논문에서는 원격 화상 회의에서 화자 간의 자연스러운 시선 맞춤(eye contact)을 위한 깊이 영상의 후처리 필터링 기술을 제안한다. 제안하는 방법은 3 차원 비디오 시스템의 핵심 기술들 가운데 하나인 깊이 영상의 탐색 기술과 영상 합성 기술을 사용해서 화자의 정면시점 영상을 합성한다. 하지만, 깊이 영상을 탐색하는 과정에서 객체의 경계 불일치, 시간적 상관도 저하 등의 문제가 발생하기 때문에, 이를 해결하기 위해 시간축으로 확장된 결합형 양방향 필터(joint bilateral filter)를 제안한다. 실험 결과를 통해, 제안하는 깊이 영상의 후처리 필터링 기술이 정면시점 합성영상의 화질을 향상시켰고, 원격의 화자와 시선 맞춤이 가능한 것을 확인했다.

Ⅰ. 서 론

3 차원 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. 기존의 원격 화상 회의 시스템은 화자의 시선과 카메라의 렌즈의 위치가 달라 화자 간 시선 불일치가 발생한다. 시선 불일치 문제는 화자끼리의 대화의 집중을 떨어뜨리고 몰입감을 떨어뜨리기 때문에 이러한 문제를 해결하기 위해 많은 연구 기관들에 의해 시선 맞춤을 위한 연구가 진행되었다. 최근, 독일의 HHI 연구소에서는 앞서 언급한 주요 기술들을 이용한 3 차원 원격 화상회의 시스템을 개발했다 [1]. 이 방법은 4 대의 카메라를 이용해서 화자의 깊이 정보를 탐색한 다음, 영상 합성 과정을 통해 원격의 화자의 시선을 맞춘다. 하지만, 이 방법은 성능에 비해 하드웨어 구성이 너무 복잡하고 시스템 구축에 너무 많은 비용이 든다는 단점이 있다.

본 논문에서는 깊이 영상의 후처리 필터링 기술을 통한 원격 화상 회의의 시선 맞춤 방법을 제안한다. 제안하는 방법은 기존의 깊이 탐색 기술에서 발생하는 객체의 경계 불일치, 시간적 상관도 저하 등의 문제 등을 해결하기 위해 [2], 깊이 영상에 대해 색상 영상의 경계 정보와 색상 정보를 깊이 영상의 필터링에 사용하는 결합형 양방향 필터(joint bilateral filter)를 사용한다 [3]. 제안하는 방법을 통해 카메라와 화자와의 거리를 계산할 수 있게 되고, 최종적으로 정면시점 영상합성 방법을 통해 화자가 정면을 바라보는 듯한 영상을 합성해서 시선 맞춤을 가능하게 한다.

Ⅱ. 제안하는 깊이 영상의 후처리 필터링 기술

깊이 탐색 기술을 이용하면 3 차원 장면의 깊이 정보를 획득할 수 있지만, 그림 1 에서 알 수 있듯이 색상 영상과 변위 영상의 객체 경계가 맞지 않는 문제점이 발생한다. 이러한 문제점은 전역적 방법을 사용했을 때 자주 발생하는 것으로 영상을 합성하는 과정에서 오차를 발생시킬 뿐만 아니라, 다양한 3 차원 응용 분야에 깊이 정보를 사용할 수 없게 한다.

(a) 색상 영상 (b) 변위 영상 (c) 겹침 영상

그림 1. 색상 영상과 변위 영상의 경계 불일치

본 논문에서는 기존 기술에서 사용한 양방향 필터를 시간축으로 확장하는 깊이 영상 후처리 필터링 방법을 제시한다. 제안하는 방법은 필터 적용 범위를 시간축으로 확장한 3 차원 필터를 사용한다. 시간축으로 확장된 결합형 양방향 필터는 다음과 같이 정의된다.

$$D(x,y,y) = \arg\min_{d \in \vec{d}_p} \frac{\sum_{u \in u_p} \sum_{v \in v_p} \sum_{w \in w_p} W(u,v,w) C(u,v,w,d)}{\sum_{u \in u_p} \sum_{v \in v_p} \sum_{w \in w_p} W(u,v,w)} \quad (1)$$

위 식에서 각 변수는 $p=(x,y)$, $d_p=\{D(x-1,y,t), D(x+1,y,t), D(x,y-1,t), D(x,y+1,t), D(x,y,t-1), D(x,y,t+1)\}$, $u_p=\{x-r,\ldots,x+r\}$, $v_p=\{y-r,\ldots,y+r\}$, $w_p=\{t-r,\ldots,t+r\}$와 같고 $D(x,y,t)$는 변위값을 나타낸다. 위 식에서 양방향 필터의 가중값인 $W(u,v,w)$와 정합 비용 $C(u,v,w,d)$는 다음과 같이 정의된다.

$$W(u,v,w)=\exp\left\{-\frac{\|I(x,y,y),I(u,v,w)\|^2}{2\sigma_R^2}\right\}\times \exp\left\{-\frac{\sqrt{(x-u)^2+(y-v)^2+(t-w)^2}}{2r^2}\right\} \quad (2)$$

$$C(u,v,w,d)=\min\{\lambda\Gamma,|D(u,v,w)-d|\} \quad (3)$$

여기서 λ 는 정합 오차 $C(u,v,w,d)$의 상한을 결정하기 위한 상수를 나타내고 L 은 스테레오 정합에 사용된 변위값의 범위를 나타낸다. 그리고 $I(x,y,t)$는 색상 영상의 화소값을 나타낸다.

위와 같은 결합형 양방향 필터를 시간축으로 확장해서 깊이 영상에 적용하게 되면 깊이 영상 내의 대부분의 영역의 시간적 상관도를 향상시킬 수 있지만 움직임이 급격한 객체에 대해서는 필터링 이후에 오히려 깊이값의 오차가 발생할 수 있다. 따라서, 본 논문에서는 이를 해결하기 위해 시간축으로 발생하는 외곽 오차를 제거하기 위한 방법을 추가적으로 적용한다.

제안하는 방법은 양방향 필터의 3 차원 윈도우 내의 이전 화면과 이후 화면에서 (x,y) 위치의 깊이값을 조사한다. 이 깊이값들 가운데 현재 화면과의 차이가 한계값 이상이고 화소값 차이 역시 또다른 한계값 이상일 경우에는 외곽 오차가 포함된 화면이라 판단하여 이 화면을 제거한 후 필터링을 수행한다. 이를 수식으로 나타내면 다음과 같다.

$$w_{outlier\_reduction}=\{w_p\,|\,|I(x,y,t)-I(x,y,w_p)|<2\lambda L,\ |D(x,y,t)-D(x,y,w_p)|<\lambda L\} \quad (3)$$

Ⅲ. 실험 결과 및 분석

본 논문에서 제안하는 시선 맞춤 방법의 성능을 평가하기 위해, 그림 2 와 같은 시스템을 구축하고 영상을 획득했다. 카메라의 간격은 42 인치 디스플레이를 기준으로 했을 때, 약 93cm, 시청거리는 2.7m 로 설정했다. 영상을 획득하기 위해서 배경의 변화는 없고 배경과 화자의 거리는 거의 차이가 없도록 제한사항을 두었다. 카메라 모델은 Point Grey Research 사의 Grasshopper 이며 해상도는 1280x960 으로 설정했다.

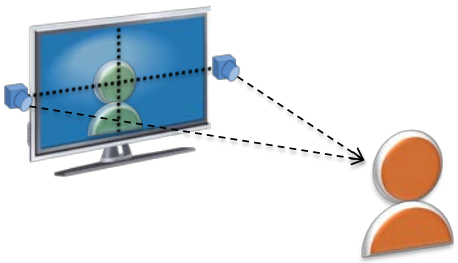

그림 2. 시선 맞춤을 위한 스테레오 카메라 시스템

그림 3 은 깊이 영상의 후처리 필터링 적용 유무에 따른 정면시점 합성영상의 화질을 보여준다. 그림 3 에서 좌측 2 개 영상은 좌시점 카메라, 우시점 카메라 각각에서 획득된 영상이며, 우측 2 개 영상은 정면시점으로 영상을 합성한 결과이다. 그림 3(b)에서 알 수 있듯이, 제안하는 깊이 영상 필터링 기술을 사용했을 때, 화자의 경계 부분이 더 자연스러운 것을 확인할 수 있었다.

(a) 필터링 적용 전

(b) 필터링 적용 후

그림 3. 깊이 영상 후처리 필터링 결과

Ⅳ. 결론

깊이 영상의 후처리 필터링 기술을 통한 원격 화상회의의 시선 맞춤 방법을 제안했다. 제안하는 방법은 깊이 영상에 대해 시간축으로 확장한 결합형 양방향 필터를 사용함으로써, 객체의 경계 불일치, 시간적 상관도 저하 등의 문제 등을 해결할 수 있었다. 결과적으로, 후처리 기술을 통해 정제된 깊이 영상을 이용해서 정면시점에서의 영상을 합성했기 때문에, 기존의 방법에 비해 보다 자연스러운 합성영상을 얻을 수 있었고, 화자 사이에 시선 맞춤이 가능한 것을 확인할 수 있었다.

ACKNOWLEDGMENT

본 연구는 방송통신위원회의 방송통신미디어 원천기술개발 사업의 일환인 “ IPTV 용 Interactive 시점제어 기술개발 [09912-03002]” 과제의 결과물입니다.

참 고 문 헌

[1] P. Kauff and 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.

[2] S. Lee, I. Shin, and Y. Ho, "Gaze-corrected View Generation using Stereo Camera System for Immersive Videoconferencing," IEEE Transactions on Consumer Electronics, vol. 57, no. 3, pp. 1033-1040, 2011.

[3] Q. Yang, L. Wang, and N. Ahuja, "A constant-space belief propagation algorithm for stereo matching," in Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, pp. 1458-1465, 2010.