

자연스러운 눈맞춤 영상을 위한 홀 채움 방법

고은상 호요성
광주과학기술원
{esko, hoyo}@gist.ac.kr

Hole Filling Method for Natural Eye Gaze Correction

Eunsang Ko Yo-Sung Ho
Gwangju Institute of Science and Technology (GIST)

요 약

영상회의 장치에서 눈맞춤 영상을 생성하기 위해 어파인 변환(affine transformation)을 이용하면 반올림 오차(round-off error) 때문에 홀이 발생한다. 이러한 홀을 채우려면 홀 영역을 가리키는 홀 채움 마스크가 필요하다. 홀 채움은 보통 홀 채움 마스크를 참조하여 홀이 아닌 이웃 화소값들을 기반으로 손상된 영상을 복원하는 작업이다. 따라서 홀 영역을 정확히 검출하고 적당한 개수의 이웃 화소값을 참조해야 자연스러운 홀 채움 영상을 생성할 수 있다. 한편, 눈맞춤 영상을 생성할 때 얼굴 특징점을 이용해 얼굴 변환 마스크를 만들고, 얼굴 변환 마스크에만 어파인 변환을 수행한다. 이 논문에서는 얼굴 특징점에도 어파인 변환을 수행하고 수정된 얼굴 변환 마스크를 획득하여 정확한 홀 채움 마스크를 구한다. 또한, 홀 채움 마스크에서 레이블링을 수행하여 큰 홀 영역을 제거한다. 마지막으로, 어파인 변환을 수행할 때 기존 영상의 좌표값을 이용하여 자연스러운 홀 채움 영상을 생성한다. 제안하는 방법으로 홀 채움을 수행한 결과, 연속적인 눈맞춤 동영상에서 이웃값들을 참조하여 홀 채움을 수행한 영상보다 자연스러움을 확인했다.

1. 서론

영상회의는 사용자간에 멀리 떨어져 있어도 같은 장소에 있는 것처럼 대화할 수 있게 한다. 최근 통신 기술이 발전하여 일반인들이 직장뿐만 아니라 집에서도 스마트폰이나 웹캠을 사용하여 영상회의를 많이 사용하고 있다. 그러나 일반적인 영상회의의 시스템은 모니터와 카메라의 위치 차이 때문에 사용자간에 눈맞춤을 어렵게 한다. 눈맞춤은 효율적인 의사소통을 제공하고 서로 간에 신뢰를 주는 방법이므로 영상회의의 시스템에서 중요하게 다뤄지는 문제이다 [1, 2].

이에 따라, 지금까지 영상회의의 시스템에서 눈맞춤을 위한 다양한 방법이 제안되어 왔다. 그러나 기존에 제안된 방법들은 복잡하고 값 비싼 하드웨어를 필요로 하거나 시스템 설치가 매우 까다롭다. 이런 문제점들을 해결하기 위해, 최근에는 영상회의의 시스템에도 3차원 영상 기술이 사용되고 있다. 특히, 여러 대의 색상 카메라와 깊이 카메라를 사용하거나 여러 대의 색상 카메라로 깊이 정보를 예측하여 색상 영상과 깊이 영상을 눈맞춤이 되는 가상 시점에 투영시켜 눈맞춤 영상을 생성하는 방법이 대표적이다 [1, 3].

그러나, 깊이 정보를 사용한 가상 시점의 눈맞춤 영상을 생성하는 방법은 깊이 정보를 예측하거나 낮은 해상도의 깊이 정보를 업샘플링(upsampling)할 때 높은 시간 복잡도를 가진다. 이 문제를 해결하기 위해, 최근에는 Kinect를 사용하여 지금까지 제안된 방법보다 간단한 설치로 눈맞춤 영상을 실시간으로 생성하는 방법이 제안되었다 [4]. Kuster 방법은 사용자의 얼굴에서 66개의 특징점을 검출하고 사용자의 얼굴에 대해서만

눈맞춤을 수행한다. 따라서 카메라 시점과 모니터 시점에서 사용자를 촬영하여 각 시점에서의 특징점 정보와 깊이 정보를 획득하는 초기화 과정이 필요하다. 이 논문에서는 Kuster의 방법을 바탕으로 Kinect v2를 사용하여, 3차원 어파인 변환(affine transformation)을 수행함으로써 Full HD 해상도의 눈맞춤 영상을 강인하게 생성하는 방법을 제안한다. 또한, 3차원 어파인 변환의 반올림 오차(round-off error)에 의해 발생하는 홀을 채우는 방법을 제안한다.

2. 눈맞춤 영상 생성 방법

제안된 눈맞춤 영상 시스템은 27인치 모니터와 Kinect v2를 사용한다. Kinect v2는 모니터 아래에 설치되며, 효율적인 눈맞춤 영상 생성을 위해 색상 카메라를 모니터 중간에 맞추고, Kinect v2는 사용자의 얼굴을 향하도록 각도를 올렸다. 키넥트와 사용자의 최적의 거리는 850 mm에서 1250 mm까지이다. 그림 1은 제안된 시스템을 보여준다.

제안하는 눈맞춤 영상 생성 방법은 Kuster 방법과 마찬가지로 카메라 시점과 모니터 시점에서 사용자의 얼굴 특징점을 획득하는 초기화 과정이 필요하다. 그림 2(a)와 2(b)는 각각 카메라 시점과 모니터 시점이다. 각 시점에서의 사용자를 볼 때, 27인치 모니터를 사용하는 영상회의의 시스템도 눈맞춤이 되지 않는 것을 볼 수 있다. 사용자의 얼굴 특징점은 이마와 눈, 광대뼈, 턱에서 11개를 뽑아 두 시점의 일치 쌍을 구해 3행 4열의 3차원 어파인 변환 행렬을 예측한다.



그림 1. 눈맞춤 영상 생성 시스템

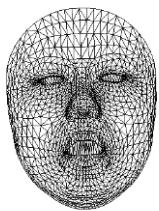


(a) 카메라 시점

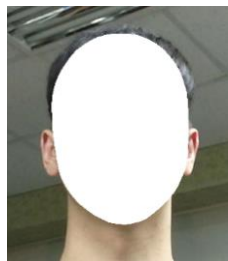
(b) 모니터 시점

그림 2. 카메라 시점과 모니터 시점에서의 얼굴 특징점

초기화 과정이 끝나면, 얼굴 특징점을 추적하면서 모든 특징점을 이용해 얼굴 변환 마스크를 생성한다. Kinect v2를 사용하여 얼굴 검출을 수행하면 1347개의 특징점과 7890개의 인덱스 정보를 획득하여 2630개의 삼각형으로 얼굴 모델을 그릴 수 있다. 얼굴 변환 마스크는 얼굴 모델이 2차원으로 투영되고 삼각형이 채워진 형태이다. 그림 3(a)와 3(b)는 각각 얼굴 모델과 색상 영상에 그린 얼굴 변환 마스크다.



(a) 3차원 얼굴 모델



(b) 2차원 마스크

그림 3. 얼굴 모델과 얼굴 변환 마스크

다음 과정은 깊이 정보를 3차원 어파인 변환에 사용하기 위해 512*424 해상도의 깊이 영상을 색상 영상의 해상도로 업 샘플링을 수행한다. 3차원 어파인 변환을 수행하기 전에 색상 좌표를 카메라 공간 좌표로 변환하는데, 카메라 공간 좌표의 Z 값은 카메라와 사용자의 거리를 미터 (meter) 단위로 표현하는 값이므로, 깊이값을 1000으로 나눈 값으로 대입한다. X 좌표는 사용자의 좌우를 나타내므로 오차를 줄이기 위해 눈맞춤 영상 x 좌표값은 기존 x 좌표값을 그대로 사용한다. Y 좌표는 색상 좌표에서 변환하기 위해 (1)을 사용하고, 다시 눈맞춤 영상의 y 좌표값으로 변환하기 위해 (2)를 사용한다. p와 f값은 각각

색상 카메라의 내부 파라미터인 주점 (principal point)과 초점거리 (focal length)다. 3차원 어파인 변환 결과, Y 값과 Z 값이 작아짐으로써 눈맞춤 영상이 생성된다. 그림 4는 3차원 어파인 변환을 수행하여 눈맞춤 영상을 생성하고 홀 채움은 수행되지 않은 결과다. 제안하는 방법은 그림 4의 결과에서 홀을 채워 눈맞춤 영상을 생성한다.

$$Y = \frac{-(y - p_y) \cdot Z}{f_y} \quad (1)$$

$$y = \frac{-Y \cdot f_y}{Z} + p_y \quad (2)$$

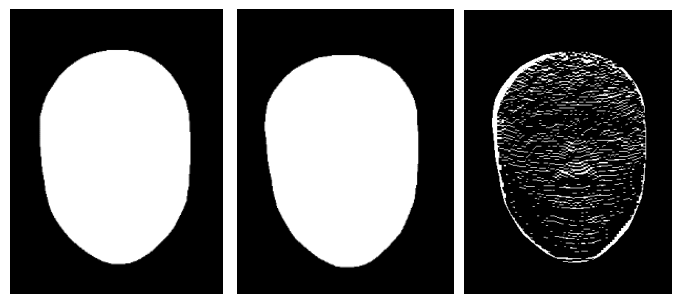


그림 4. 홀 채움이 되지 않은 눈맞춤 영상

3. 제안하는 홀 채움 방법

3.1 홀 채움 마스크 획득

홀 채움을 수행하기 위해서는 홀 영역을 가리키는 홀 채움 마스크를 필요로 한다. 홀 채움은 보통 홀이 아닌 이웃 화소값들을 이용하여 홀을 채워나가는 작업이다. 따라서 홀이 아닌 영역에 대해 홀 채움을 수행하거나 홀 영역을 제대로 검출하지 못하면 부자연스러운 결과를 얻는다. 제안하는 홀 채움 방법은 정확한 홀 탐지를 위해 얼굴 특징점에도 어파인 변환을 수행하고, 수정된 특징점을 가지고 수정된 얼굴 변환 마스크를 획득한다. 수정된 얼굴 변환 마스크는 눈맞춤 영상에서의 얼굴 영역과 일치하므로, 마스크에서 눈맞춤이 되는 화소값들을 제외시키면 홀 채움 마스크가 된다. 그림 5(a)와 5(b)는 각각 얼굴 변환 마스크의 원본과 수정된 것이고, 그림 5(c)는 그림 4의 홀 채움 마스크를 보여준다.

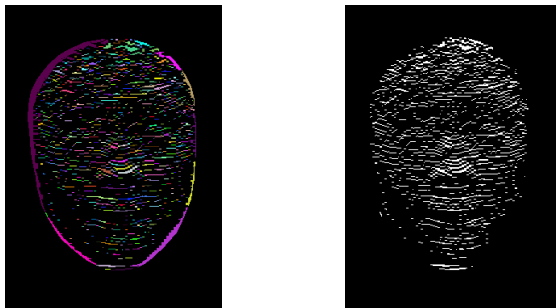


(a) 원본 마스크 (b) 수정된 마스크 (c) 홀 채움 마스크

그림 5. 얼굴 변환 마스크와 홀 채움 마스크

3.2 홀 채움 마스크에서 큰 홀 영역 제거

그림 5(c)를 볼 때, 얼굴의 경계 부분에서는 깊이 정보의 오차 등의 이유로 여러 개의 큰 홀이 생기는 것을 볼 수 있다. 그림 5(c)의 마스크를 사용하여 홀 채움을 수행한다면, 얼굴의 경계 부분일 뿐만 아니라 홀에서 먼 영역의 화소값을 참조해야 하므로 부자연스러운 홀 채움 결과를 얻는다. 따라서 제안하는 방법은 홀 채움 마스크에서 홀들을 레이블링(labeling)하고, 덩어리 된 홀에서 화소값의 개수를 정렬하여 일정 개수 이상인 덩어리 홀을 제거하는 알고리즘을 수행한다. 레이블링은 연결 요소 (connected component) 방법을 사용하여, 큰 홀 영역과 얼굴 내의 작은 홀 영역이 한 덩어리로 묶이지 않게 4방향성 탐색 방법을 사용했다 [5]. 그림 6(a)는 홀 채움 마스크에서 서로 다른 홀 영역을 색상으로 표현하고, 그림 6(b)는 여러 개의 큰 홀 영역이 제거된 홀 채움 마스크다.



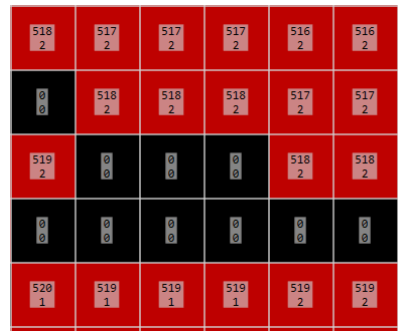
(a) 마스크 레이블링 (b) 최종 홀 채움 마스크

그림 6. 홀 채움 마스크에서 큰 영역 제거

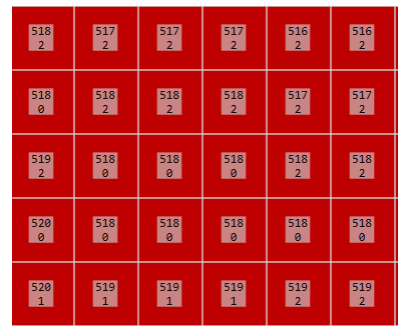
3.3 홀 채움 수행

보통 홀 채움을 수행할 때는 Telea 방법을 많이 사용한다. Telea 방법은 fast marching method (FMM) 알고리즘을 기반으로 동작한다. FMM 알고리즘은 Eikonal 편미분방정식을 푸는 수치 계산법 중에 하나로 다익스트라 (Dijkstra) 알고리즘과 유사하다. 이를 기반으로 하는 Telea 방법은 홀 영역의 바깥 경계에서 시작하여 안쪽으로 점차 진행하며 이웃 화소값들의 가중치 합으로 홀 위치에 화소값을 생성한다.

그러나, 눈맞춤 영상의 홀을 볼 때 반올림 오차로 인해 작은 영역의 많은 홀들이 존재한다. 작은 영역의 홀을 채울 때는 Telea 방법이 비효율적이므로, 제안하는 홀 채움 방법은 눈맞춤 영상을 생성할 때 저장한 원본 색상 영상의 y 좌표값을 참조하여 홀 채움을 수행한다. 또한, y 좌표값을 저장할 때, 눈맞춤 영상의 y 좌표값의 소수점에 따라 홀 채움을 수행할 때 화소값 참조 위치를 선택한다. 그림 7은 y 좌표값을 저장한 영상의 일부를 보여준다. 두 번째 원소값에 따라 화소값 참조 위치를 위와 아래에서 선택한다. 참조 위치가 선택되면, 기존 색상 영상에서 참조 위치의 화소값을 가져와 눈맞춤 영상에 홀 채움을 수행하고, y 좌표 참조 영상에도 그 값을 채운다. 그러나 두 번째 값은 0을 유지하여 다른 홀에서 홀 채움을 수행할 때는 그 값을 참조하지 않는다. 제안하는 방법으로 홀 채움을 수행한 뒤에도 큰 영역의 홀에서 홀 채움이 되지 않았다면, Telea 방법을 사용하여 최종 눈맞춤 영상을 생성한다 [6].



(a) 홀 채움 수행 전



(b) 홀 채움 수행 후

그림 7. 기존 영상의 y 좌표 참조 영상

4. 실험결과

그림 8(a)는 눈맞춤을 수행하기 전 원본 색상 영상이고, 3차원 어파인 변환을 수행하면 그림 4와 같은 홀이 있는 눈맞춤 영상을 획득한다. 그림 8(b)와 그림 8(c)는 각각 Telea 방법과 제안하는 방법으로 홀 채움을 수행한 눈맞춤 영상의 최종 결과이다. 언뜻 보기에는 두 결과가 서로 차이가 없어 보이지만, 영상을 확대했을 때 눈동자에서 홀 채움 결과가 미미하게 다를 수 있다. 그림 9는 눈맞춤 영상의 최종 결과에서 눈 부분을 4배 확대하여 두 결과의 차이를 보여준다. 제안하는 방법으로 홀 채움을 수행한 결과가 Telea 방법보다 눈의 검은자와 흰자의 영역을 잘 보존했다. 비록 작은 차이일지라도, 연속적인 눈맞춤 동영상을 볼 때는 이 차이는 뚜렷해진다. 수행 시간 차이에서는 제안하는 홀 채움 방법이 평균 0.08초 빠르게 수행되어 실시간 눈맞춤 영상 생성에 기여했다.



(a) 원본 색상 영상

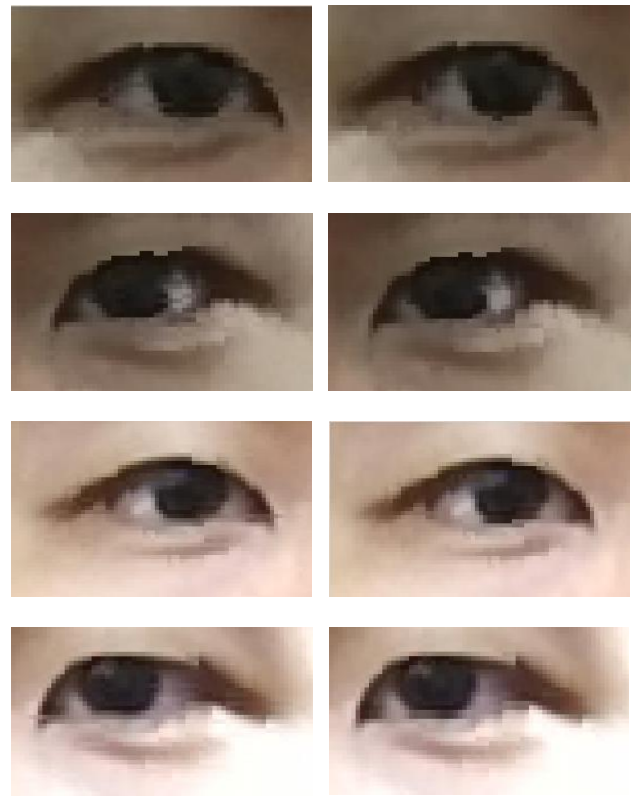


(b) Telea 홀 채움 결과



(c) 제안하는 홀 채움 결과

그림 8. 눈맞춤 영상의 최종 결과



(a) Telea 방법

(b) 제안하는 방법

그림 9. 홀 채움 수행 결과 차이 영상

5. 결론

이 논문에서는 Kinect v2를 사용하여 사용자의 얼굴 특징 점을 검출하고 3차원 어파인 변환을 수행하여 눈맞춤 영상을 생성했다. 또한, 눈맞춤 영상에서 3차원 어파인 변환의 반올림 오차 때문에 발생하는 홀 채움을 수행했다. 제안하는 방법은 얼굴 특징점에도 3차원 어파인 변환을 수행하여 정확하게 홀을 탐지할 수 있는 홀 채움 마스크를 생성하여 정확한 홀 채움을 수행했다. 두 번째로, 얼굴의 경계 부분에서 깊이 정보의 부정확함으로 인해 큰 홀 영역이 발생하는데, 홀 채움 마스크에서 레이블링을 수행하고 큰 홀 영역을 제거하여 자연스러운 홀 채움 결과를 생성했다. 세 번째로, 더 빠르고 자연스러운 홀 채움 영상을 획득하기 위해 기존 색상 영상의 화소값의 참조 위치를 선택하는 홀 채움 방법을 수행했다. 실험결과로는 Telea 방법과는 크게 차이가 없었지만 눈동자에서 보다 깨끗한 결과를 얻어 실시간으로 생성하는 눈맞춤 동영상에서 자연스러운 결과를 얻었다.

감사의 글

이 논문은 2015년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2011-0030079)

참고문헌

- [1] S. B. Lee, I.Y. Shin, and Y.S 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.
- [2] E. Ko, W.S. Jang, and Y.S. Ho, "Eye Gaze Correction for Video Conferencing Using Kinect v2," Pacific-Rim Conference on Multimedia (PCM), Part II, LNCS 9315, pp. 571-578, 2015.
- [3] L. Zhang and T. Wa James, "Stereoscopic Image Generation Based on Depth Images for 3D TV," IEEE Transactions on Broadcasting, vol. 51, no. 2, pp. 191-199, 2005.
- [4] C. Kuster, T. Popa, J.C. Bazin, C. Gotsman, and M. Gross, "Gaze Correction for Home Video Conferencing," ACM Transactions on Graphics, vol. 31, no. 6, pp. 1-6, 2012.
- [5] K. Wu, E. Otoo, and K. Suzuki, "Two Strategies to Speed up Connected Component Labeling Algorithms," LBNL Tech Report LBNL-59102, 2005.
- [6] A. Telea, "An Image Inpainting Technique Based on the Fast Marching Method", Journal of Graphics Tools, vol. 9, no. 1, pp. 25-36, 2004.