

2대의 Kinect 카메라를 이용한 3차원 물체의 복원 구현

신동원 호요성

광주과학기술원 영상통신연구실

{dongwonshin, hoyo}@gist.ac.kr

Implementation of 3D Object Reconstruction using a Pair of Kinect Cameras

Dong-Won Shin Yo-Sung Ho

Gwangju Institute of Science and Technology (GIST)

요약

본 논문에서는 2대의 Kinect 카메라를 이용하여 실세계의 3차원 객체에 대한 복원을 수행하는 방법을 제안한다. 먼저 깊이 가중치가 추가된 계층적 결합형 양방향 필터를 이용하여 Kinect로부터 얻은 원본 깊이 영상을 보정한다. 그리고 카메라 캘리브레이션을 이용하여 카메라의 내부 파라미터와 외부 파라미터를 획득한다. 이를 이용해 3차원 워핑을 수행하여 각 시점의 데이터를 3차원 공간에 점군 모델로 복원하고 표면 모델링 방법을 이용하여 3차원 객체의 매끄러운 표면 모델을 생성한다. 실시간에 가까운 속도를 내기 위해서 계층적 결합형 양방향 필터와 3차원 워핑을 병렬 처리 프레임워크인 CUDA로 구현하여 고속화하였다. 실험을 통해 분리된 각 시점에서의 깊이 정보를 하나의 통합된 3차원 공간에 복원할 수 있었고 초당 5 fps의 속도로 동작하는 것을 확인하였다.

1. 서론

우리가 생활하는 실제 세계는 X, Y, Z 축으로 만들어진 3차원으로 구성되어 있다. 3차원 복원이란 이러한 실세계의 3차원 공간 속에 존재하는 물체를 컴퓨터와 같은 가상환경 속에 재현하는 것을 말한다. 최근에 3차원 복원은 게임, 영화, 광고, 건축, 측량, 예술과 같은 다양한 분야에서 많은 각광을 받고 있다.

구체적인 예로서 3차원 얼굴 복원을 들 수가 있다. 사람의 얼굴은 인간의 몸에서 가장 많은 근육을 가지고 있는 부분이다. 우리는 이렇게 수많은 근육들을 바탕으로 약 7000가지의 표정을 지을 수 있는데 이를 통해 의사소통과 감정전달을 더욱 더 효과적으로 할 수 있게 만들어준다. 우리는 3차원 복원을 활용해 자신의 얼굴을 가진 캐릭터를 만들어 직접 캐릭터 상품을 제작할 수 있고, 최근에 많은 성장을 하고 있는 모바일 콘텐츠 산업에도 활용할 수 있다. 또한 성형 및 미용 의료 분야에서는 3차원 복원된 얼굴을 통해 성형 전 후 환자의 모습을 짧은 시간에 3차원으로 복원해 상담에 이용하려는 시도가 늘고 있다. 또한 3차원 인체 복원을 통해 사람의 몸을 가상 세계에 복원하여 가상의 의류를 체험하게 하고 실제 제품의 구매로 이어지도록 유도하는 실시간 의류 착용 서비스도 개발되고 있다.

이러한 3차원 복원 기술은 인간에 대한 정보뿐만 아니라 객체의 복원에도 널리 쓰일 수 있다. 그 예로 들 수 있는 것이 문화재 복원이다. 지난 2008년 승례문 화재 사건이 발생 당시 미리 스캔해 놓았던 승례문의 3차원 정보가 승례문을 복원 시키는데 큰 도움이 되었다고 한다. 문화재청은 승례문 화재 사건 이후 3차원 가상 복원의 중요성을 절감하고 주요 문화재의 3차원 촬영을 진행하였다. 또한 문화재 복원에서 끝나지 않고 화재 수사에도 활용이 되었는데 사진이나 동영상으로

사건 현장을 보존하던 기존방식에서 더 나아가 3차원 복원 작업을 통해 화재 현장을 기록함으로써 화재수사에 대해 큰 도움을 주었다 [1].

3차원 복원을 위한 3차원 정보의 획득 방법에는 다양한 것들이 있지만 주로 능동형 방식과 수동형 방식으로 나뉜다. 능동형 방식은 적외선의 패턴을 물체에 사영하여 패턴의 왜곡 정도를 계산한 다음 3차원 물체의 깊이 정보를 획득하거나 레이저를 물체에 직접 비추어 되돌아오는 시간을 측정하여 깊이를 측정하는 방법 등 주로 직접 물체에 광선을 쏘아서 물체의 3차원 정보를 획득하는 방법이다. 이러한 능동형 방식은 정확도가 상대적으로 높지만 카메라 센서의 가격이 비싸고 사용이 어렵다는 단점이 있다. 수동형 방식은 두 시점 이상의 색상 영상을 촬영한 뒤에 색상 영상 사이의 영상 패턴 매칭을 통해 좌우 시차를 얻어내고 이로부터 각각의 시점에 대한 3차원 깊이 정보를 획득하는 방법이다. 이러한 수동형 방식은 상대적으로 저가인 색상 카메라만을 이용하여 시스템을 구성할 수 있으나, 영상의 해상도가 낮고, 처리에 많은 시간이 소요된다는 단점이 존재한다.

최근에는 능동형 방식에서는 광학 소자 기술이 많이 발달하고, 능동 센서의 가격이 저렴해지는 한편, 수동형 방식에서는 컴퓨팅 성능의 증가와 병렬 처리 기술의 발달로 인하여 전반적인 3차원 복원에 대한 연구가 활발하게 진행되고 있다 [2].

3차원 복원에서 주요 쟁점으로 고려되는 것은 첫 번째로 다양한 시점의 카메라간의 캘리브레이션이 정확하게 수행되어야 한다는 것이다. 캘리브레이션을 통해 각 카메라의 내부, 외부 파라미터를 얻고 이 파라미터를 이용하여 각 시점에서 얻은 색상과 깊이 정보를 통일된 3차원 공간에 정확하게 재현해 낼 수가 있어야 한다. 다음으로 방대한 양의 3차원 정보를 짧은 시간 내에 빠르게 처리할 수 있는 알고리즘이 개발되어야 한다. 최근에는 컴퓨터 계산 성능의 향상과 병렬 처리 기술

의 확산이 영상 처리 알고리즘의 가속화에 많은 도움을 주고 있다. 마지막으로 각 시점에서 얻어지는 깊이 영상에 대한 정확도가 높아야 한다. 앞서 설명했던 3차원 깊이 획득 방법으로 얻은 원본 깊이 영상은 실제 3차원의 세계를 표현하기에 상대적으로 정확성이 떨어진다. 센서에러, 폐색영역, 경계불일치 등의 원인으로 깊이 값이 갖지 않는 부분이 존재할 수 있으며, 색상 영상에 비해 상대적으로 낮은 해상도를 가지는 깊이 영상은 3차원 복원 처리를 위해 색상 영상과 같은 해상도로 일치시키는 과정도 필요하다.

본 논문은 이러한 주요 쟁점들을 고려하면서 2대의 Kinect 카메라를 이용하여 3차원 복원을 하는 방법에 대해 설명한다. 2장에서는 제안하는 3차원 복원 시스템의 물리적인 구조를 설명하고 3장에서는 제안하는 3차원 복원 알고리즘의 흐름을 설명한다. 4장에서는 이러한 3차원 복원 시스템을 이용한 3차원 복원의 결과를 살펴본 뒤 5장에서는 살펴본 내용들을 종합하여 결론과 향후과제를 논의한다.

2. 시스템 세팅

본 논문에서 제안하는 방법은 2대의 Kinect 카메라를 수렴형으로 설치하여 3차원 물체를 복원하는 방법에 대한 것이다. Kinect는 Microsoft에서 출시된 3차원 비전용 카메라로서 저렴한 가격에 비해 성능이 좋은 카메라를 가지고 있고 두터운 사용자층을 가지고 있다. Kinect는 RGB 카메라와 적외선 프로젝터, 적외선 카메라, 소리를 감지할 수 있는 마이크를 내장하고 있다. Structured light 방식을 사용하는 Kinect는 내장된 적외선 프로젝터에서 임의의 순서로 나열된 적외선 도트 패턴을 물체에 사영하면 사영된 패턴은 Kinect로부터 물체의 거리에 따라 왜곡된 형태로 형성되는데 이 왜곡된 패턴을 적외선 카메라로 촬영한 후 삼각측량 방법을 이용해 깊이를 측정하게 되는 원리이다 [3].

본 논문에서는 이러한 Kinect 카메라를 그림 1과 같이 물체의 전면에 위치시키고 45도 정도로 물체 방향으로 회전하여 수렴형으로 2대를 설치한다.

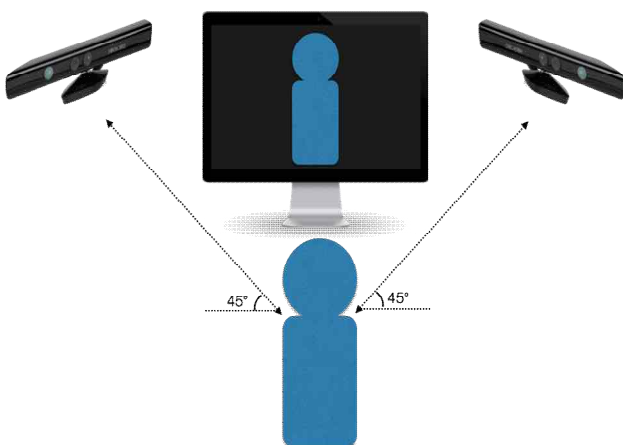


그림 1. 제안하는 시스템의 물리적 구조

그 다음 물체를 두 Kinect의 수렴위치에 존재하도록 위치시킨다.

복수의 Kinect를 하나의 컴퓨터에 연결하기 위해서는 메인보드에 직접적으로 연결된 USB 허브 하나에 Kinect 하나가 연결되어야 동작한다는 단점이 있다. 따라서 Kinect 개수에 비례하여 USB 허브를 메인보드에 추가적으로 장착시켜 주어야 한다.

3. 3차원 복원

제안하는 3차원 복원 방법의 흐름도는 그림 2와 같다.

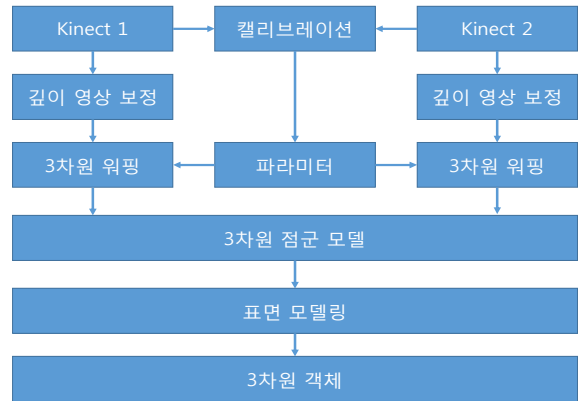


그림 2. 제안하는 3차원 복원 방법의 흐름도

먼저 Kinect 카메라의 내부, 외부 파라미터를 얻기 위해 캘리브레이션 과정을 오프라인으로 수행한다. 그런 다음 각각의 Kinect 카메라로부터 색상 영상과 깊이 영상을 받은 뒤에 깊이 영상에 대해 홀 부분을 채우는 보정 작업을 수행한다. 다음으로 앞에서 획득한 카메라의 내부, 외부 파라미터를 이용하여 3차원 워핑을 수행해 분리된 각 시점에서의 깊이 정보를 통합된 3차원 공간으로 보내준다. 이 결과로 실제 물체에 대한 3차원 점군 모델을 3차원 공간에 생성되는데 이 데이터에 표면 모델링을 수행하면 매끄러운 표면을 가진 3차원 객체를 얻을 수 있다.

3. 1 캘리브레이션

복수의 카메라를 이용하여 3차원 복원을 하기 위해서는 캘리브레이션 과정이 반드시 필요하다. 본 논문에서는 일반적으로 많이 사용되는 카메라 캘리브레이션 방법인 평면의 체크무늬 패턴을 사용하여 캘리브레이션 영상을 찍고 Matlab calibration toolbox를 이용하여 카메라의 내부 파라미터와 외부 파라미터를 얻는 방법을 이용하였다 [4]. 체크무늬 패턴은 직사각형 하나의 크기가 30*30 mm이고 가로로 7개 세로로 5개의 패턴을 캘리브레이션에 이용하였다. 그림 3은 제안된 카메라 시스템에서 촬영된 평면의 체크무늬 패턴 영상이다.

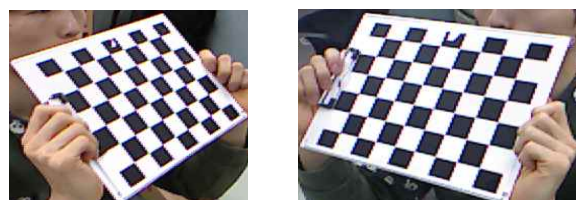


그림 3. 카메라 캘리브레이션을 위한 패턴 영상

2대의 Kinect에서 하나의 장면을 동시에 촬영해야 하며 캘리브레이션 하고자 하는 패턴의 특징점이 영상에 모두 나타나야 한다. 각각의 Kinect에서 10개의 패턴 영상을 찍으면 총 20개의 패턴영상을 얻게 된다. 이 패턴영상을 Matlab calibration toolbox를 통해 캘리브레이션 작업을 거치게 되면 각각의 카메라에 대해서 내부 파라미터와 외부 파라미터를 얻을 수가 있게 된다. 이 내부, 외부 파라미터를 이용하여 3차원 워핑을 수행하면 분리된 각각의 카메라에서 얻은 깊이 정보를 통합된 하나의 3차원 공간에 표현할 수 있게 된다.

3. 2 깊이 영상 보정

Kinect로부터 얻은 원본 깊이 영상은 센서 에러, 폐색 영역 등으로 인해 깊이 값이 가지지 않은 부분을 많이 포함하고 있다. 따라서 이러한 부분을 정확한 깊이 값으로 채워주기 위해 많은 방법들이 연구되고 있다. 본 논문에서는 3차원 복원 시스템의 실시간성을 고려하여 병렬 처리 프로그래밍을 이용하여 구현된 계층적 결합형 양방향 필터를 사용하였다 [5]. 또한 기존의 결합형 양방향 필터 연산에서는 픽셀간의 색상 차이와 거리 차이만을 고려하여 수행되었으나 본 논문에서는 픽셀간의 깊이 값 차이도 결과에 반영을 하였다. 수식 (1)은 기존의 결합형 양방향 필터의 수식을 나타낸다.

$$D_o(x, y) = \frac{\sum_{u \in U_p} \sum_{v \in V_p} W(u, v) D_i(u, v)}{\sum_{u \in U_p} \sum_{v \in V_p} W(u, v)} \quad (1)$$

여기서 D_o 는 출력 깊이 영상을 나타내고 (x, y) 는 필터 커널에서의 중심 화소의 위치를 나타낸다. D_i 는 입력 깊이 영상이고 (u, v) 는 필터 커널에서의 인접 화소의 위치를 나타낸다. 그리고 W 는 가중치 인자를 나타낸다. 본 논문에서는 가중치인자 W 에 들어가는 식에서 깊이 값 차이를 결과에 반영하기 위해 W 를 수식 (2)와 같이 수정하였다.

$$W(u, v) = \begin{cases} 0 & D_i(u, v) = 0 \\ g(u, v) \cdot f(u, v) \cdot d(u, v) & otherwise \end{cases} \quad (2)$$

$g(u, v)$ 는 중심 화소와 인접 화소 간의 거리 차이를 반영하는 Gaussian filter의 가중치 인자이고, $f(u, v)$ 는 색상 차이를 반영하는 Euclidean filter의 가중치 인자이다. 여기 까지가 기존의 결합형 양방향 필터의 가중치 인자이고 여기에 중심화소와 인접화소 간의 깊이 값 차이를 반영하는 가중치인자 $d(u, v)$ 를 추가하였다. 수식 (3)은 가중치 인자 $d(u, v)$ 의 수식을 나타낸다.

$$d(u, v) = \begin{cases} 1, & |D_i(x, y) - D_i(u, v)| \leq th \\ 0, & |D_i(x, y) - D_i(u, v)| > th \end{cases} \quad (3)$$

$d(u, v)$ 에서는 깊이 영상에서 중심 화소와 인접 화소간의 깊이 값 차이의 절대 값이 임계값을 넘게 되면 0을 반환하고 임계값 내부에 존재하면 1을 반환하여 깊이 값 차이가 많이 나면 결과에 반영하지 않도록 하였다.

3. 3 3차원 워핑

다음으로 색상 영상과 보정된 깊이 영상을 이용하여 3차원 워핑을 수행한다. 3차원 워핑은 각각의 카메라에서 얻은 내부 파라미터와 외부 파라미터를 이용하여 각 시점의 영상을 통합된 3차원 공간상으로 보내어 나타내는 알고리즘이다. 이를 통해 우리는 합성된 3차원 점군 모델을 얻을 수 있다. 수식 (4)는 3차원 워핑에 대한 식을 나타낸다 [6].

$$M_w = R^{-1} A^{-1} \begin{bmatrix} d \cdot x \\ d \cdot y \\ d \end{bmatrix} - R^{-1} t \quad (4)$$

R 은 카메라의 외부 파라미터 중에 회전 행렬을 의미하고 A 는 내부 파라미터를 의미한다. 그리고 x 와 y 는 영상에서의 좌표를 의미하고 d 는 깊이 영상에서 (x, y) 위치의 깊이 값을 나타낸다. 마지막으로 t 는 외부 파라미터의 이동 행렬을 나타낸다. 이 수식을 이용해서 우리는 3차원의 점 좌표 x, y, z 값을 가진 벡터 M_w 를 얻게 된다.

3차원 워핑을 통해 얻은 3차원 좌표를 이용하여 객체를 공간상에 나타내면 그림 4와 같다.



그림 4. 3차원 워핑을 수행하여 얻은 결과

3차원 공간상에 객체를 나타내었기 때문에 객체에 대한 시점을 자유자재로 변경하여 관찰할 수 있지만 이 모델은 점들이 모여서 형성된 점군 모델이기 때문에 점과 점 사이를 채우지 못하고 비어있는 공간이 보이게 된다. 이러한 문제를 해결하기 위해서 3차원 점군 모델을 표면 모델로 변환하는 과정이 필요하다. 그림 5는 표면 모델링을 수행하는 방법을 나타낸다.

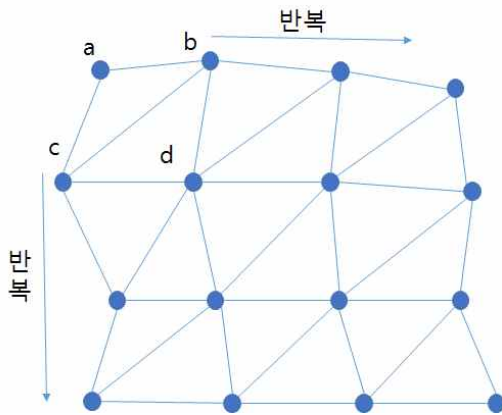


그림 5. 표면 모델링 수행 방법

점군 모델에서 기준이 되는 점 a의 우측에 있는 점b와 하단에 위치한 점c를 이용하여 삼각형을 만든다. 그리고 기준이 되는 점 a의 우하단에 위치한 점d를 이용하여 삼각형 abc와 빗변을 공유하는 삼각형을 만든다. 이러한 방법을 이용해 반복적으로 표면을 생성해 나간다. 그런데 이런 방식으로 표면을 만들게 되면 기준이 되는 점과 멀리 떨어진 점에 대해서도 연결을 하는 문제가 발생할 수 있기 때문에 임계값을 넘지 않는 범위 내에서 연결하여 표면을 생성하도록 하였다.

4. 실험 결과

본 논문에서 연산에 사용한 CPU는 Intel Xeon 2.53Ghz이고 GPU는 Nvidia Geforce GTX Titan을 사용하였다. 계층적 결합형 양방향 필터에서 사용한 인자 값들에 대해 Gaussian filter의 sigma값은 15, Euclidean filter의 sigma값은 100, 필터의 반경은 15, 계층의 개수는 5개로 하였다. 또한 표면 모델링에서 거리에 따른 임계값을 10으로 설정하고 3차원 복원을 수행하였다. 그림 6은 제안하는 3차원 복원방법을 통해 생성된 3차원 모델을 나타낸 그림이다.



그림 6. 제안하는 3차원 복원 방법의 결과

앞서 살펴보았던 그림 4와 비교하여 상대적으로 많은 부분의 홀이 채워진 것을 확인할 수 있다. 그림 7은 객체에 대해 다양한 시점에서 획득한 그림이다. 여러 각도에서 객체를 볼 수 있으며 또한 평균 5 fps의 속도로 동작하는 것을 확인하였다.

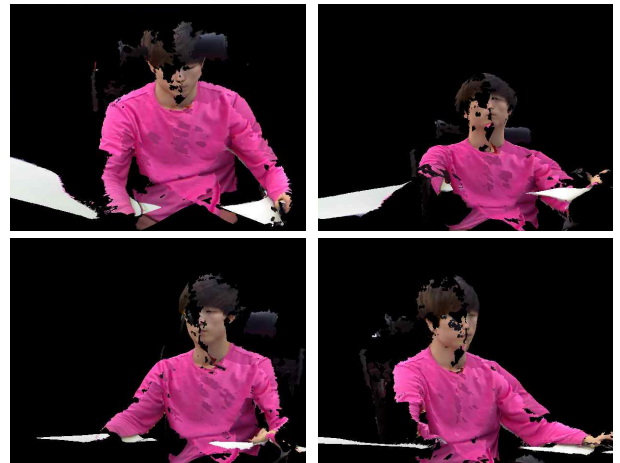


그림 7. 3차원 모델을 다양한 시점에서 획득한 사진

5. 결론 및 향후과제

본 논문에서는 2대의 Kinect를 이용하여 3차원 복원을 수행하는 방법에 대해서 살펴보았다. 2대의 Kinect를 물체의 전방에 수렴형으로 위치시키고 캘리브레이션을 통해 카메라의 내부, 외부 파라미터를 얻은 뒤 깊이 가중치가 추가된 계층적 결합형 양방향 필터 수식을 이용하여 원본 깊이 영상을 보정한 다음 3차원 워핑을 통해 각 시점의 깊이 영상을 하나의 통합된 3차원 공간에 표현하였다. 이렇게 얻어진 객체에 대한 점군 모델을 표면 모델링을 통해 매끄러운 표면으로 구성하여 보다 사실감 있는 3차원 객체를 복원하였다. 하지만 아직 완벽하게 깊이 값이 채워진 모델을 얻지는 못하여 깊이 영상 보정과 3차원 워핑에서 정확한 깊이 값을 얻는 방법에 대해 연구해야 할 것이고 초당 30 fps의 실시간 속도에 접근하기 위해 많은 연구가 필요할 것이다.

감사의 글

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

참고 문헌

- [1] http://biz.chosun.com/site/data/html_dir/2013/03/25/2013032502165.html
- [2] 윤승욱, 황본우, 김갑기, 임성재, 최진성, 구분기, "3D 얼굴 복원 기술 분석 및 연구 동향," 전자통신동향분석, 27권, 3호, pp 12-21, 2012. 06.
- [3] 장주용, 류문욱, 박순찬, "거리영상 기반 동작인식 기술동향," 전자통신동향분석, 29권, 1호, pp 11-20, 2014. 02.
- [4] Camera Calibration Toolbox for MATLAB: <http://www.vision.caltech.edu/bouguetj>.
- [5] D. Shin, Y. Ho, "Real-time Depth Image Refinement using Hierarchical Joint Bilateral Filter," Journal of The Korean Society of Broadcast Engineers, pp. 140-147, Mar. 2014
- [6] W. R. Mark, L. McMillan, and G. Bishop, "Post-rendering 3D Warping," in Proc. of Symposium on Interactive 3D Graphics, pp. 7-16, Apr. 1997.