

특집논문 (Special Paper)

방송공학회논문지 제26권 제3호, 2021년 5월 (JBE Vol. 26, No. 3, May 2021)

<https://doi.org/10.5909/JBE.2021.26.3.247>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

3D 스켈레톤을 이용한 3D 포인트 클라우드의 캘리브레이션

박 병 서^{a)}, 강 지 원^{a)}, 이 솔^{a)}, 박 정 탁^{a)}, 최 장 환^{a)}, 김 동 욱^{a)}, 서 영 호^{a)†}

A New Calibration of 3D Point Cloud using 3D Skeleton

Byung-Seo Park^{a)}, Ji-Won Kang^{a)}, Sol Lee^{a)}, Jung-Tak Park^{a)}, Jang-Hwan Choi^{a)},
Dong-Wook Kim^{a)}, and Young-Ho Seo^{a)†}

요 약

본 논문에서는 3D(dimensional) 스켈레톤을 이용하여 다시점 RGB-D 카메라를 캘리브레이션 하는 새로운 기법을 제안하고자 한다. 다시점 카메라를 캘리브레이션 하기 위해서는 일관성 있는 특징점이 필요하다. 또한 높은 정확도의 캘리브레이션 결과를 얻기 위해서는 정확한 특징점의 획득이 필요하다. 우리는 다시점 카메라를 캘리브레이션 하기 위한 특징점으로 사람의 스켈레톤을 사용한다. 사람의 스켈레톤은 최신의 자세 추정(pose estimation) 알고리즘들을 이용하여 쉽게 구할 수 있게 되었다. 우리는 자세 추정 알고리즘을 통해서 획득된 3D 스켈레톤의 관절 좌표를 특징점으로 사용하는 RGB-D 기반의 캘리브레이션 알고리즘을 제안한다. 다시점 카메라에 촬영된 인체 정보는 불완전할 수 있기 때문에, 이를 통해 획득된 영상 정보를 바탕으로 예측된 스켈레톤은 불완전할 수 있다. 불완전한 다수의 스켈레톤을 효율적으로 하나의 스켈레톤으로 통합한 후에, 통합된 스켈레톤을 이용하여 카메라 변환 행렬을 구함으로써 다시점 카메라들을 캘리브레이션 할 수 있다. 캘리브레이션의 정확도를 높이기 위해서 시간적인 반복을 통해서 다수의 스켈레톤을 최적화에 이용한다. 우리는 실험을 통해서 불완전한 다수의 스켈레톤을 이용하여 다시점 카메라를 캘리브레이션 할 수 있음을 증명한다.

Abstract

This paper proposes a new technique for calibrating a multi-view RGB-D camera using a 3D (dimensional) skeleton. In order to calibrate a multi-view camera, consistent feature points are required. In addition, it is necessary to acquire accurate feature points in order to obtain a high-accuracy calibration result. We use the human skeleton as a feature point to calibrate a multi-view camera. The human skeleton can be easily obtained using state-of-the-art pose estimation algorithms. We propose an RGB-D-based calibration algorithm that uses the joint coordinates of the 3D skeleton obtained through the posture estimation algorithm as a feature point. Since the human body information captured by the multi-view camera may be incomplete, the skeleton predicted based on the image information acquired through it may be incomplete. After efficiently integrating a large number of incomplete skeletons into one skeleton, multi-view cameras can be calibrated by using the integrated skeleton to obtain a camera transformation matrix. In order to increase the accuracy of the calibration, multiple skeletons are used for optimization through temporal iterations. We demonstrate through experiments that a multi-view camera can be calibrated using a large number of incomplete skeletons.

Keyword : point cloud, 3d reconstruction, rgb-d, skeleton, pose estimation

I. 서론

최근에 RGB와 depth 센서가 결합된 RGB-D 센서 (카메라)가 보편화 되었고, 이것은 다양한 분야에서 널리 사용되고 있다. RGB-D 카메라는 객체의 형태와 주위 환경의 3차원 구조를 비교적 정확하고 빠르게 추출할 수 있게 도와준다. RGB-D 카메라는 SLAM과 navigation^{[1][2]}, tracking^{[3][4]}, object recognition과 localization^[5], pose estimation^[6] 그리고 3D model reconstruction^[7] 등과 같은 다양한 분야를 급속히 발전시켜왔다. RGB-D 카메라에서 컬러 정보는 전형적으로 RGB 카메라를 이용하여 획득된다. 반면에 depth 정보는 time-of-flight (ToF) 카메라, laser range scanner, 그리고 structured-light (SL) sensor와 같은 다양한 방식을 이용하여 획득된다^[8]. robotics 및 vision 분야에서 널리 사용되는 RGB-D 카메라에는 Microsoft Kinect 시리즈^{[9][10]}, the Asus Xtion^[11], 그리고 Intel RealSense^[12] 등이 있다. RGB-D 카메라를 이용하여 신뢰성 있고 정확한 scene representation을 위해서, 각 카메라의 intrinsic calibration과 두 센서 사이의 extrinsic calibration이 필요하다. 최근에는 이들을 위한 정확한 parameter set들이 미리 주어지고, 이 값들은 디바이스 내부의 non-volatile memory에 저장되어 있다. Real-time scanning/integration 및 3D geometry model의 capturing과 같은 다수의 RGB-D 카메라를 이용하여 촬영을 수행하는 응용에서는 다수의 카메라들 사이의 extrinsic calibration이 매우 중요하다^[13]. RGB-D 카메라는 RGB 정보 뿐만 아니라 depth 정보가 동시에 취득되기 때문에, 다시점 RGB-D 카메라들 사이의 calibration은 RGB

정보만을 이용한 다시점 카메라 기반의 calibration과 같은 고전적인 방식과 달리 depth 정보를 함께 이용한다. RGB 정보와 depth 정보를 동시에 이용하여 calibration을 수행할 때, depth 정보의 생성 방식과 과정을 고려한다^[14].

정확한 카메라 파라미터를 획득하기 위해서 다양한 연구들이 수행되어 왔다. 크게는 structured light 기반의 depth sensing 기법^[15-18]과 ToF 카메라 기반의 depth sensing을 위한 캘리브레이션 기법^[19-22]으로 분류할 수 있다. 성공적인 캘리브레이션을 위한 잡음 제거, 패턴 생성, 센서 품질, depth error 예측 및 보정, 그리고 thermal and environmental distortion^[23] 등을 고려한다면, 캘리브레이션 기술의 범위를 명확히 정의하는 것은 다소 어려운 일이다. 캘리브레이션을 하기 위해서 체스보드를 사용하는 방식과 사용하지 않고 영상의 특징점을 이용하는 방식 등이 있다. 만일 다시점 카메라 시스템을 설치한 이후, 카메라의 물리적인 움직임으로 인해 다시 캘리브레이션을 해야 하는 것은 매우 번거로운 일이다. 체스보드를 사용하는 방식의 경우에 2차원 혹은 3차원 체스보드를 가져와서 이동하면서 영상을 촬영한 후에 캘리브레이션을 수행해야 한다. 객체를 촬영한 후에 특징점을 찾고, 이 특징점을 이용하여 캘리브레이션을 하는 방식은 체스보드가 필요 없는 장점은 있으나 정확하고 일관성 있는 특징점을 찾아야 하는 어려움이 존재하고, 특징점을 구하고자 하는 대상물의 종류에 따라서 결과가 달라지는 특성을 갖는다. 본 논문에서는 사람의 스켈레톤을 구한 후에 이것을 특징점으로 사용하여 다시점 카메라를 캘리브레이션 하는 새로운 알고리즘을 제안하고자 한다^[24].

본 논문은 다음과 같이 구성된다. 먼저 2장에서는 스켈레톤들 간의 정합에 사용하는 최적화 함수 기반의 카메라 변환 행렬을 구하는 방법을 설명한다. 3장에서는 캘리브레이션 알고리즘을 제안한다. 4장에서는 실험 결과를 보이고 5장에서 우리의 논문을 마무리한다.

II. RGB-D 기반의 정합

본 절에서는 스켈레톤의 3차원 관절들을 이용하여 3D 스켈레톤의 정합을 위한 좌표계 변환 파라미터를 획득하는 방법에 대해 설명한다^[25]. 3D 스켈레톤의 좌표계를 월드 좌

a) 광운대학교 전자재료공학과(Kwangwoon university Electronic Materials Engineering)

✉ Corresponding Author : 서영호(Young-Ho Seo)
E-mail: yhseo@kw.ac.kr
Tel: +82-940-8362

ORCID: <http://orcid.org/0000-0003-1046-395X>

※ 이 논문은 2021년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(NRF-2018R1D1A1B0704322013).

이 논문은 2021년도 광운대학교 대학혁신지원사업에 의해 연구되었음.

※ This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education in 2021 (NRF-2018R1D1A1B07043220). The present Research has been conducted by the Research Grant of Kwangwoon University in 2021.

• Manuscript received April 6, 2021; Revised May 14, 2021; Accepted May 14, 2021.

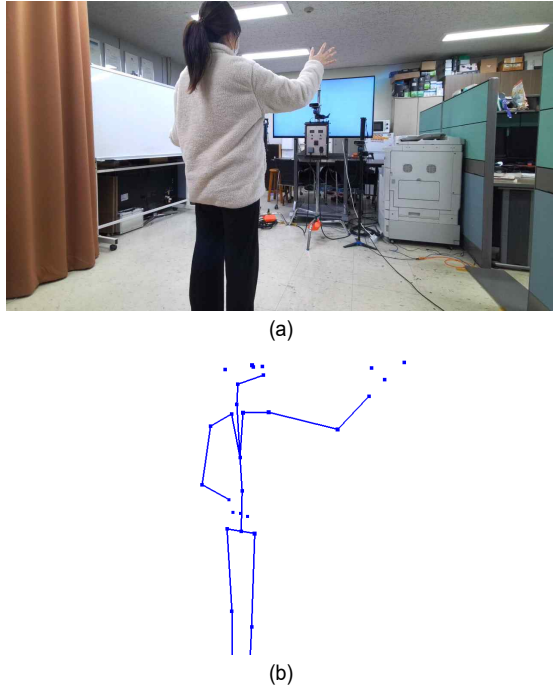


그림 1. 실험에 사용될 영상 및 3D 스켈레톤 (a) 한 시점에서 촬영된 영상, (b) 영상의 스켈레톤
Fig. 1. Captured image and 3D skeleton for experiment (a) a captured image in a view-point, (b) the estimated skeleton

표계로 설정하여 카메라 외부 파라미터를 구하고, 특징점 매칭을 이용하여 포인트 클라우드를 정합한다. 그림 1은 우리가 사용할 영상과 그로부터 예측한 3D 스켈레톤 정보이다.

월드 좌표계를 구한 뒤에 카메라 좌표계에서 월드 좌표계로의 변환 행렬을 구해야 한다. 서로 다른 카메라에서 동일한 사람에 대한 스켈레톤이 획득되면, 두 카메라 사이의 같은 월드 좌표계를 계산할 수 있다. 같은 월드 좌표계를 두 카메라가 공유하면, 두 카메라 좌표계 사이의 변환 관계가 유도된다. 깊이 영상의 픽셀좌표계에서 3차원 공간상의 좌표계로 변환한 후에, RGB 영상의 픽셀좌표계로 re-projection하는 방법을 이용하여 깊이 영상에서 RGB영상으로의 캘리브레이션이 수행된다. 또한 3차원 공간의 좌표계로 변환하기 위해 깊이영상의 픽셀좌표계를 정규 픽셀 좌표계로 변환한다. 다음으로 실제 3차원 공간상에서 RGB 카메라와 깊이 카메라 사이의 변환관계에 대한 외부 파라미터를 이용하여 정규 픽셀 좌표계를 RGB 카메라의 3차원 좌

표계로 변환한다. 마지막으로 3차원 RGB 카메라 좌표계를 깊이정보로 나누어 정규 픽셀 좌표계로 변환한 후에, RGB 카메라의 내부좌표계를 곱해줌으로써 RGB 영상의 픽셀 좌표계로 변환한다. 3차원 좌표를 8대의 모든 카메라에서 획득하고 좌표들을 매칭 시킨다. 매칭 된 좌표끼리 같은 위치에 위치하도록 하는 좌표변환 파라미터를 추정하는 방법을 사용한다. 우리는 최적화 알고리즘을 사용하여 파라미터들 간의 오차 함수를 만든 후에, 경사 하강법을 이용하여 오차 함수의 해를 구한다^[26].

좌표계 변환행렬에는 x, y, z 축 각각의 회전각, 평행이동 값 그리고 스케일링 팩터 까지 총 7개의 파라미터가 포함되어 있다. 기준 카메라 좌표계로 구하고자 하는 시점의 좌표계를 변환하는 파라미터를 구한다. 이 과정은 식 (1)로 정의된다. 깊이 영상에서 발생하는 잡음 성분 및 카메라 제작 오차에 의해 촬영된 깊이 값이 실제거리와 다른 값이 출력될 수 있기 때문에, 이것을 보완하기 위한 파라미터로 스케일링 팩터를 새롭게 도입한다. X_{ref} 는 기준 카메라의 좌표계이고, X_i 는 추정하는 카메라의 좌표계이다. $R_{i \rightarrow ref}$, $t_{i \rightarrow ref}$ 그리고 $S_{i \rightarrow ref}$ 은 각각 기준 카메라 좌표계로의 회전 변환 행렬, 평행이동 행렬, 그리고 스케일링 factor이다. 초기에 $R_{i \rightarrow ref}$ 은 단위행렬로 설정하고, $S_{i \rightarrow ref}$ 과 $t_{i \rightarrow ref}$ 는 각각 1과 0으로 설정한다^[26].

$$X'_i = R_{i \rightarrow ref} S_{i \rightarrow ref} X_i + t_{i \rightarrow ref} \quad (1)$$

오차 함수 f_{Error} 는 X_{ref} 와 X'_i 의 유클리드 제곱 거리 (Squared Euclidean Distance, SED)의 평균값이고, 식 (2)와 같이 정의된다. 식 (2)에서 $X'_i(j)$ 가 $X_{ref}(j)$ 에 근접하게 되면 f_{Error} 가 영으로 수렴하게 된다. 즉, $X_i(j)$ 을 $X_{ref}(j)$ 로 변환하기 위한 최적의 $R_{i \rightarrow ref}$, $t_{i \rightarrow ref}$ 그리고 $S_{i \rightarrow ref}$ 를 추정하였다는 의미가 된다.

$$f_{Error} = \frac{1}{N} \sum_{j=0}^N \| X_{ref}(j) - X'_i(j) \|_2^2 \quad (2)$$

이 함수를 모든 좌표계 변환 파라미터에 대해 편미분한 후에 오차 함수값이 최소가 되도록 파라미터를 갱신해 나

가는 과정을 식 (3)으로 정의한다. 식 (3)은 n 번째 파라미터 P_n 과 f_{Error} 의 편미분 결과를 이용하여 P_{n+1} 를 갱신하는 과정을 나타낸다. α 는 학습률을 나타내는 상수이다.

$$P_{n+1} = P_n - \alpha \frac{\partial f_{Error}}{\partial P_n} \quad (3)$$

III. 제안한 캘리브레이션 알고리즘

이번 장에서는 다시점 카메라 시스템에서 3D 스켈레톤을 이용하여 포인트 클라우드 기반의 공간을 구성할 수 있는 캘리브레이션 알고리즘을 제안한다.

1. 3D 공간 생성

스켈레톤 세트들에서 매칭 되는 관절(joint)의 좌표를 이용하여 각 카메라들의 외부 파라미터들을 구하는 방법을 자세히 설명하여 보자. 스켈레톤을 이용하여 캘리브레이션

을 수행하기 위해서 우리는 다수의 프레임 세트를 이용한다. 먼저, 여러 프레임 세트 중에서 하나의 프레임 세트를 선택한다. 이 프레임 세트에서 하나의 카메라로부터 획득된 스켈레톤을 참조 스켈레톤으로 선택하고, 나머지 스켈레톤 중에서 하나를 타겟 스켈레톤으로 선택한다. 두 개의 스켈레톤을 선택할 때에는 참조 관절을 포함하여 최소한 3개 이상의 조인트가 대응이 되는 조건을 만족해야 한다. 선택된 두 개의 스켈레톤은 기본 관절을 기준으로 정렬한다. 이를 통해서 두 카메라 사이의 초기 변환 파라미터가 획득된다. 이러한 과정을 도식적으로 그림 2에 나타내었다.

정렬되지 않은 다시점 카메라 시스템은 임의의 위치에 분포하기 때문에 객체의 크기 혹은 위치에 따라가 각 카메라에서 획득된 영상에서는 객체의 전부 혹은 일부가 포함될 수 있다. 최악의 경우에는 모든 카메라에서 객체의 일부만을 촬영할 수도 있다. 이와 같은 환경에서 3D 체적에 대한 3D 자세 추정(pose estimation)을 위해서는 모든 카메라의 정보를 통합하여 하나의 3D 스켈레톤을 구해야만 한다.

다시점 카메라의 경우에 객체가 이동함에 따라서 각 카메라가 촬영할 수 있는 객체의 부분이 달라진다. 이 경우에 각 카메라들이 촬영한 부분 정보를 하나로 모아서 하나의 정보를 생성하는 알고리즘이 필요하다. 여기에서는 캘리브레이션을 목표로 하고 있으므로 각 위치에서 촬영한 객체의 일부분에 대한 스켈레톤 일부를 얻고, 이들을 이용하여 카메라들 사이의 변환 파라미터를 획득한다. 이를 통해서 우리는 다시 3D 복원을 수행하여 3D 공간을 재구성할 수도 있고, 3D 모델을 획득할 수도 있다.

앞에서 설명한 캘리브레이션 알고리즘에서 스켈레톤을 이용하여 카메라 좌표 변환 행렬을 구하는 개념도를 그림 3에 나타내었다. 각 카메라 좌표계를 기준으로 획득된 스켈레톤은 공간상에서 정렬되지 않는다. 2장에서 설명한 방법을 이용하여 두 카메라 사이의 좌표 변환 행렬을 구하면 그림 3(b)의 결과를 얻는다. 즉, 두 개의 카메라에서 예측된 스켈레톤을 정합하는 과정에서 두 카메라 사이의 좌표변환 행렬이 획득되고, 하나의 공통된 월드 좌표계를 기준으로 두 개의 카메라를 정렬할 수 있다.

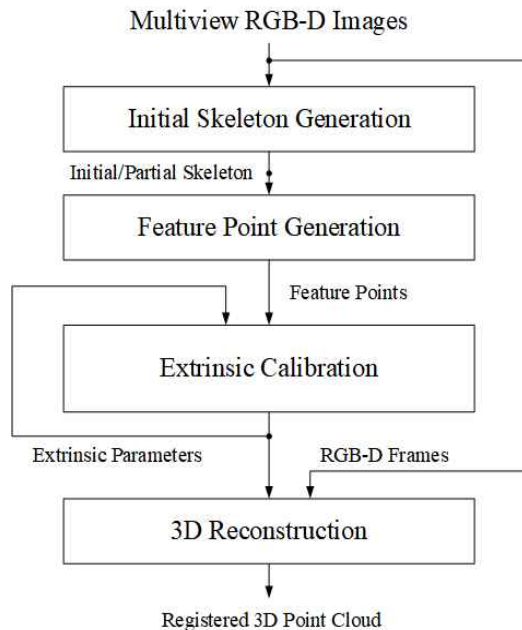


그림 2. 3D 공간 생성 알고리즘

Fig. 2. 3D space generation algorithm

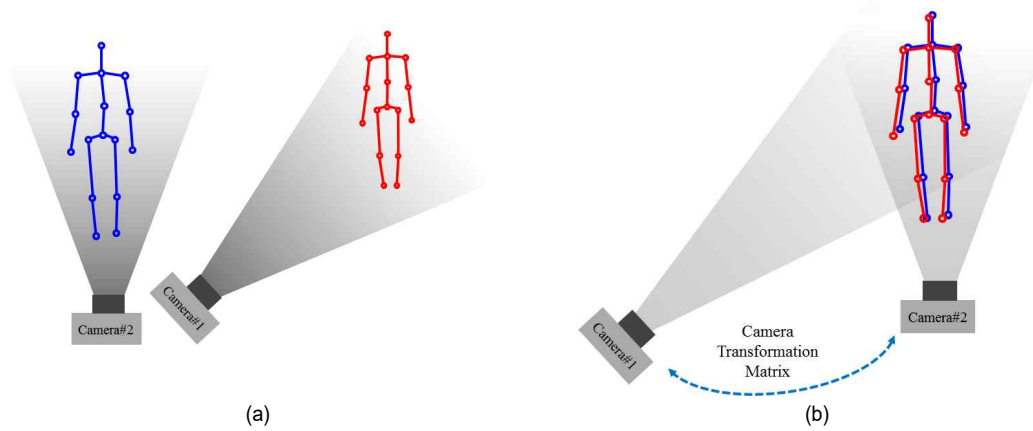


그림 3. 스켈레톤의 선택 및 베이스 조인트 정렬을 통한 초기 파라미터 생성
Fig. 3. Initial parameter generation by selecting the skeleton and aligning the base joint

2. 외부 캘리브레이션

본 절에서는 그림 4의 전체 알고리즘 중에서 외부 캘리브레이션(Extrinsic Calibration)에 대해서 설명한다. 본 논문에서는 RGB-D 센서로 Asure Kinect를 사용하고, 각 카메라별로 3D 스켈레톤을 예측하는 것은 Asure Kinect의 SDK를 이용하였다. 그러나 각 카메라별로 3D 위치 추정 알고리즘은 어떠한 것을 사용하더라도 상관없고, 카메라별로 다른 알고리즘을 사용하더라도 상관없다. 우리는 하나의 카메라로 획득한 영상에 대한 3D 위치 추정이 아니고, 불완전한 3D 스켈레톤을 완전한 하나의 스켈레톤으로 생성하고 보정하는 것에 초점을 맞추고 있다. 우리가 제안하는 외부 캘리브레이션 알고리즘을 그림 4에 플로우 차트를 이용하여 정리하였다.

불완전한 부분 스켈레톤들 중에서 하나의 참조 스켈레톤을 먼저 선택한다. 참조 스켈레톤의 선택이 전체적인 성능에 큰 영향을 주지는 않지만, 일반적으로 가장 많은 관절을 포함한 스켈레톤 중에서 하나를 참조 스켈레톤으로 선택한다. 그리고 다음으로 타겟 스켈레톤을 선택하는데, 타겟 스켈레톤은 참조 스켈레톤의 관절과 많은 관절이 중복되는 순서로 선택된다. 이 두 개의 스켈레톤은 먼저 기본 관절(척추 및 골반)을 기준으로 정렬된다. 다음으로 2장에서 설명한 최적화 함수를 기반으로 두 개의 스켈레톤을 정합하

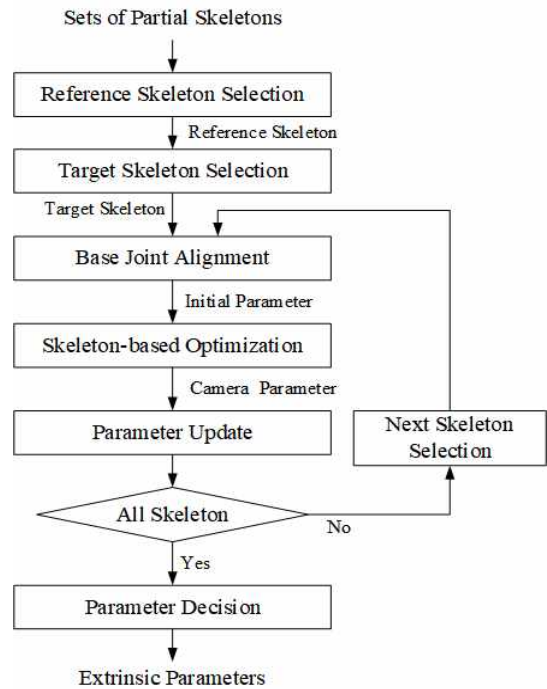


그림 4. 스켈레톤 기반의 외부 캘리브레이션을 위한 절차
Fig. 4. Proposed skeleton-based extrinsic calibration

면서 카메라의 좌표변환 파라미터를 구한다. 이 과정은 모든 스켈레톤에 대해서 반복적으로 수행되고, 많은 프레임에 대해서 일정한 오차 값으로 수렴할 때까지 이 과

정을 반복적으로 수행한다. 최종적으로 오차가 수렴하면 그때의 파라미터를 최종적인 좌표변환 파라미터로 결정한다.

IV. 실험 결과

본 장에서는 제안한 스켈레톤 기반의 3D 공간 구성을 위한 캘리브레이션 기법에 대한 실험결과를 설명한다. 먼저 실험환경을 소개하고, 다음으로 각 카메라에 대한 캘리브레이션에 대한 정량적 결과를 보인다. 마지막으로 캘리브레이션을 통해 얻은 카메라 행렬을 이용하여 재구성한 3D 공간을 보임으로써 제안한 방식의 효용성에 대해서 증명한다.

1. 실험환경

그림 5에 실험 환경에 대한 사진을 나타내었다. 실험을 위해서 총 8대의 ToF(time of flight) 방식의 RGB-D(depth) 센서인 Azure Kinect를 8대 사용하였다. 카메라의 구성은 다양할 수가 있는데, 우리는 2면에 4대씩 카메라를 배치 하여 실험을 수행하였다. 8대의 카메라는 광케이블 방식의 USB 3.0 인터페이스를 통해서 1대의 워크스테이션으로 입



(a)



(b)

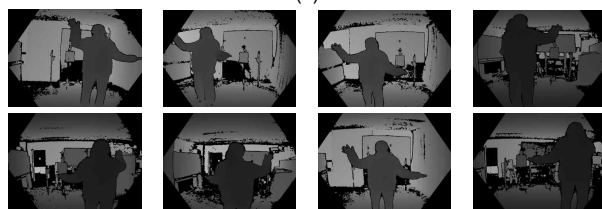
그림 5. 실험 환경 (a) 촬영, (b) 카메라 시스템

Fig. 5. Experimental environment (a) shooting, (b) camera system

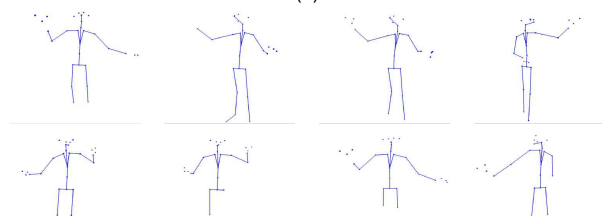
력되고, 모든 동작은 실시간(초당 30fps)으로 처리가 가능하였다. 각 카메라를 통해서 촬영된 결과 영상들과 추출된 스켈레톤을 그림 6에 정리하였다. 그림 6(a)에 보이는 것과 같이 객체(사람)의 위치에 따라서 각 카메라들은 사람의 일부분만 촬영할 수 있고, 그림 6(c)와 같이 일부분에 대한 스켈레톤만 추출이 가능하다. 제안한 방식은 이와 같이 부분적인 정보만 갖는 불완전한 스켈레톤의 관절을 특징점으로 사용하여 캘리브레이션을 수행한다.



(a)



(b)



(c)

그림 6. 다시점 카메라로부터 획득된 8개의 스켈레톤 결과 (a) RGB 영상, (b) 깊이지도, (c) 스켈레톤, (d) 카메라별 포인트 클라우드

Fig. 6. 8 skeletons acquired from multi-view cameras (a) RGB image, (b) depth map, (c) skeleton, (d) point cloud for each camera

2. 정량적 평가

8대의 시점에서 획득된 스켈레톤의 관절에 대한 최적화 결과를 그림 7에 나타내었다. 한번 최적화가 종료되면 그 다음부터는 실시간으로 3D 포인트 클라우드를 이용하여 공간의 합성이 가능하다. 1번 시점(참조)과 2번 시점(목표)의 최적화 과정을 고려하면, 1번 시점과 2번 시점의 중복된 관절 정보를 공간상에서 3차원 좌표로 겹칠 수 있도록 2번

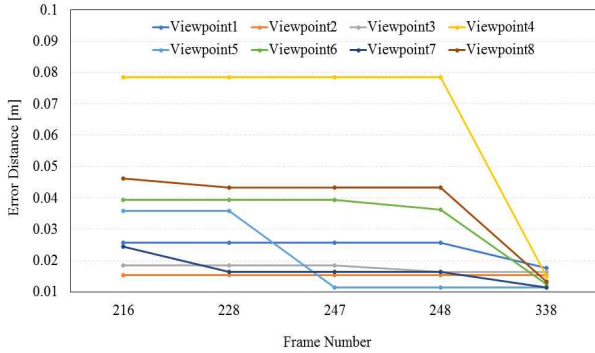


그림 7. 스켈레톤을 이용한 카메라 파라미터 최적화 그래프

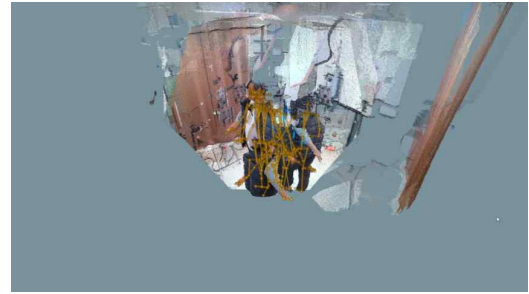
Fig. 7. Optimization graph of camera parameter using the joints of the skeletons

시점의 관절을 2번 시점으로 이동시킬 수 있는 행렬을 구한다. 카메라 행렬을 이용하여 2번 시점의 관절을 1번 시점으로 옮긴다 할지라도 두 관절 정보는 3차원 좌표계를 기준으로 위치적인 오차를 갖는다. 우리는 이 오차가 최소화되도록 카메라 행렬의 파라미터들을 최소화시키기 위한 최적화 과정을 수행하였다. 제한한 알고리즘은 다수의 프레임에 대해서 카메라 파라미터를 획득하기 위한 최적화를 수행한다. 최적화를 통해 획득된 카메라 행렬에 의한 두 시점 간의 변환 결과에 대한 오차가 일정한 값으로 수렴이 되어서 더 이상 낮은 오차가 나오지 않으면 최적화를 중단하도록 실험을 진행하였다. 아래의 실험에서는 최대 338프레임(약 11분)에 걸쳐서 최적화가 진행되었다. 실험을 통해서 우리는 약 1cm에서 2cm 이하의 오차로 수렴시킬 수 있다는 것을 확인하였다. 그림 7의 x축에 해당하는 프레임 숫자는 크게 업데이트가 되어서 오차가 비교적 크게 변화한 프레임만 나타낸 것이다.

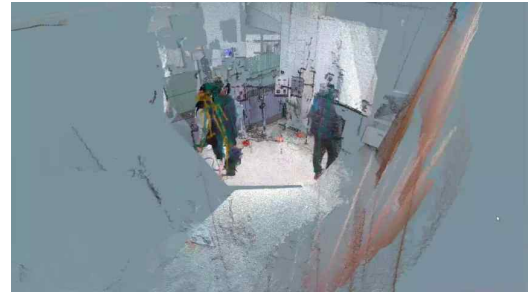
3. 시각적 평가

그림 8에는 위의 과정을 통해서 공간이 구성되는 절차를 영상 결과를 통해서 시각적으로 나타내었다. 그림 8(a)에서는 캘리브레이션 이전의 정합되지 않은 무작위적 위치의 결과이고, 그림 8(b)와 (c)는 캘리브레이션 과정의 결과이다. 그림 8(d)는 캘리브레이션 이후의 결과에 해당한다.

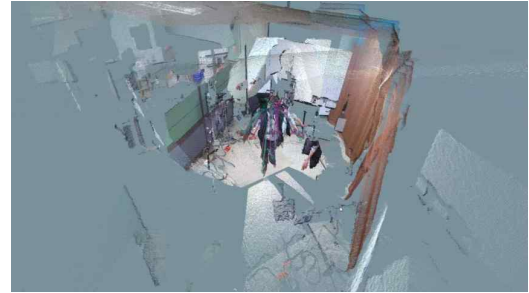
그림 9는 캘리브레이션 이후에 정합된 스켈레톤과 공간의 결과를 동시에 나타내었다. 8개의 스켈레톤은 공간 상에



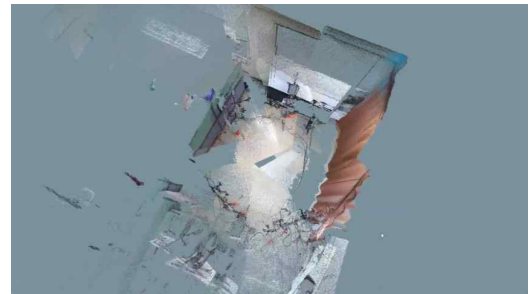
(a)



(b)



(c)



(d)

그림 8. 프레임별 스켈레톤 기반의 포인트 클라우드 최적화 영상. 프레임에 따라서 포인트 클라우드가 합쳐지면서 공간이 재구성되는 결과. (a) 초기에 정합되지 않은 결과, (b),(c) 중간 정합 결과, (d) 정합 이후에 공간이 구성된 결과

Fig. 8. Skeleton-based point cloud optimization. The result of the space reconstructed as the point clouds. (a) initially unmatched results, (b), (c) the intermediate calibrated results, and (d) the reconstructed space after calibration



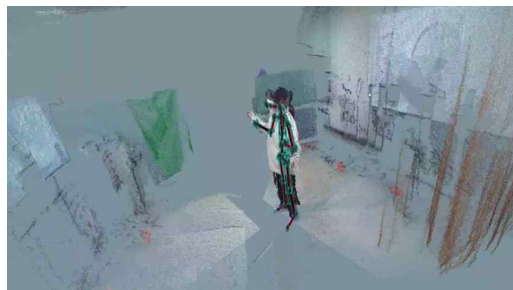
(a)



(b)



(c)



(d)

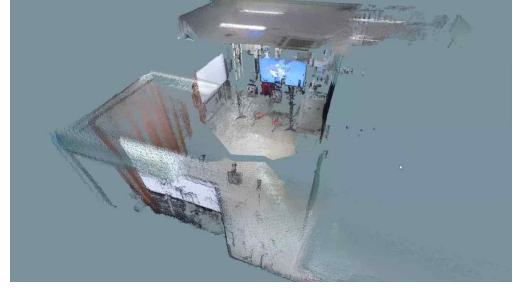
그림 9. 스켈레톤의 관절을 이용하여 카메라 정합 이후에 통합된 3D 스켈레톤의 결과 (a) 시점1, (b) 시점2, (c) 시점3, (d) 시점4

Fig. 9. Results of the integrated 3D skeleton after calibration using the joints of the skeleton; view-point (a) 1, (b) 2, (c) 3 and (d) 4

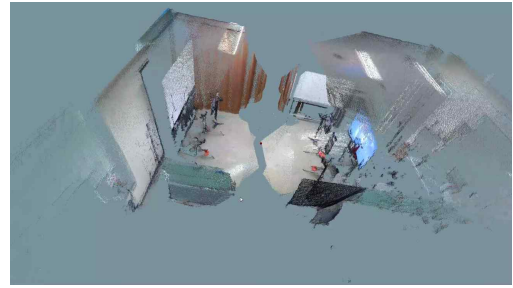
서 평균을 취하여 하나로 통합하였다. 그림 9에서 4개의 시점에서 공간과 스켈레톤을 캡처한 영상에 해당한다.



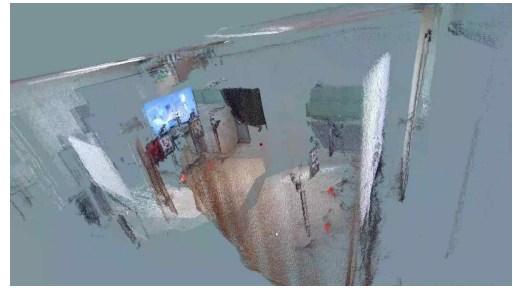
(a)



(b)



(c)



(d)

그림 10. 시점에 다른 정합된 공간에 대한 포인트 클라우드 결과 (a) 캘리브레이션 이전, (b) 시점1, (c) 시점2, (d) 시점3

Fig. 10. Point cloud results for the registered space with different view-points (a) 3D space before calibration, view-point (b) 1, (c) 2 and (d) 3 after calibration

그림 10에는 캘리브레이션을 통해서 획득된 최종적인 공간에 대한 결과 영상을 나타내었다. 그림 10(a)의 공간은

캘리브레이션 이전의 결과로써 8개의 카메라에 의한 영상이 겹쳐져서 공간상에 배치되어 있다. 이외의 영상에서는 캘리브레이션을 수행한 후에 8대의 카메라에 의해 획득된 고공간상에 정렬된 3D 포인트 클라우드 공간을 3개의 시점에서 촬영한 것이다.

V. 결 론

본 논문에서는 공간상에 다수개의 카메라가 위치해 있을 때 사람이 지나다니기만 하면 자동으로 캘리브레이션 되고, 자동으로 3D 공간이 구성될 수 있는 알고리즘에 대해 제안하였다. 사람의 스켈레톤은 비교적 일관성 있고 일정하게 추정될 수 있다는 것을 이용한다. 즉, 각 카메라에서 획득된 스켈레톤 사이의 공간상에서의 위치를 일치시키는 과정에서 카메라들 사이의 카메라 변환 행렬을 획득하였다. 이를 통해 공통된 월드 좌표계로 모든 카메라를 위치시킬 수 있었고, 3D 공간을 3D 포인트 클라우드를 이용하여 표현할 수 있었다. 카메라들 사이의 관계를 구하는 과정은 최적화 함수를 이용하였고, 그 결과로 약 1cm에서 2cm 사이의 오차를 갖도록 관절의 좌표를 대응시킬 수 있었다. 결과 영상을 살펴보면 제안한 방식을 통해서 공간을 비교적 정확하게 구성할 수 있다는 것을 증명하였다. 우리는 특별한 체크보드 혹은 차르코 보드를 사용하지 않고도 사람의 존재만으로 3D 포인트 클라우드를 성공적으로 정합하여 3D 공간을 구성할 수 있음을 보였다.

참 고 문 헌 (References)

- [1] F. Endres, J. Hess, J. Sturm, D. Cremers, and W. Burgard, "3D mapping with an RGB-D camera," *IEEE Transactions on Robotics (T-RO)*, vol. 30, no. 1, pp. 177 - 187, 2013.
- [2] M. Labb and F. Michaud, "Online global loop closure detection for large-scale multi-session graph-based SLAM," in *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2014, pp. 2661 - 2666.
- [3] M. Munaro and E. Menegatti, "Fast RGB-D People Tracking for Service Robots," *Autonomous Robots*, vol. 37, pp. 227 - 242, 2014.
- [4] C. Choi and H. Christensen, "RGB-D object tracking: A particle filter approach on GPU," in *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Tokyo, Japan, pp. 1084 - 1091, 2013.
- [5] J. Tang, S. Miller, A. Singh, and P. Abbeel, "A textured object recognition pipeline for color and depth image data," in *Proc. of the IEEE International Conference on Robotics and Automation*, Saint Paul, USA, pp. 3467-3474, 2012.
- [6] T. Munez, Y. Jembre, H. Weldegebruel, L. Chen, C. Huang and C. Yang, "The Progress of Human Pose Estimation: A Survey and Taxonomy of Models Applied in 2D Human Pose Estimation," *IEEE Access*, vol. 8, pp. 133330-133348, 2020.
- [7] M. Zollhöfer, P. Stotko, A. Görnitz, C. Theobalt, M. Nießner, R. Klein and A. Kolb, "State of the Art on 3D Reconstruction with RGB-D Cameras," *Computer Graphics Forum*, vol. 37, pp. 625-652, 2018.
- [8] Giancola S., Valenti M., Sala R. "State-of-the-Art Devices Comparison. In: A Survey on 3D Cameras: Metrological Comparison of Time-of-Flight, Structured-Light and Active Stereoscopy Technologies," *Springer Briefs in Computer Science*, pp. 29-39, 2018.
- [9] G. Chen, G. Cui, Z. Jin, F. Wu and X. Chen, "Accurate Intrinsic and Extrinsic Calibration of RGB-D Cameras With GP-Based Depth Correction," *IEEE Sensors Journal*, vol. 19, no. 7, pp. 2685-2694, 1 2019.
- [10] W. Yun and J. Kim, "3D Modeling and WebVR Implementation using Azure Kinect, Open3D, and Three.js," *2020 International Conference on Information and Communication Technology Convergence (ICTC)*, Jeju Island, pp. 240-243, 2020.
- [11] S. Yeh, Y. Chiou, H. Chang, W. Hsu, S. Liu and F. Tsai, "Performance improvement of offline phase for indoor positioning systems using Asus Xtion and smartphone sensors," *Journal of Communications and Networks*, vol. 18, no. 5, pp. 837-845, October 2016.
- [12] A. Zablatani et al., "Intel® RealSense™ SR300 Coded Light Depth Camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 10, pp. 2333-2345, 1 Oct. 2020.
- [13] G. Unal, A. Yezzi, S. Soatto and G. Slabaugh, "A Variational Approach to Problems in Calibration of Multiple Cameras," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 8, pp. 1322-1338, Aug. 2007, <https://doi.org/10.1109/TPAMI.2007.1035>.
- [14] F. Basso, E. Menegatti and A. Pretto, "Robust Intrinsic and Extrinsic Calibration of RGB-D Cameras," *IEEE Transactions on Robotics*, vol. 34, no. 5, pp. 1315-1332, Oct. 2018.
- [15] K. Khoshelham and S. O. Elberink, "Accuracy and resolution of kinect depth data for indoor mapping applications," *Sensors*, vol. 12, no. 2, pp. 1437 - 1454, 2012.
- [16] I. Mikhelson, P. Lee, A. Sahakian, Y. Wu, and A. Katsaggelos, "Automatic, fast, online calibration between depth and color cameras," *Journal of Visual Communication and Image Representation*, vol. 25, 2014.
- [17] I. V. Mikhelson, P. G. Lee, A. V. Sahakian, Y. Wu, and A. K. Katsaggelos, "Automatic, fast, online calibration between depth and color cameras," *Journal of Visual Communication and Image Representation*, vol. 25, 2014.
- [18] A. N. Staranowicz, G. R. Brown, F. Morbidi, and G. L. Mariottini, "Practical and accurate calibration of RGB-D cameras using spheres," *Computer Vision and Image Understanding*, vol. 137, pp. 102 - 114,

- 2015.
- [19] K. Zheng, Y. Chen, F. Wu, and X. Chen, "A general batch-calibration framework of service robots," in *Proc. Int. Conf. Intell. Robot. Appl. Springer*, pp. 275 - 286, 2017.
- [20] M. Lindner, I. Schiller, A. Kolb, and R. Koch, "Time-of-flight sensor calibration for accurate range sensing," *Computer Vision and Image Understanding*, vol. 114, no. 12, pp. 1318-1328, 2010.
- [21] A. Kuznetsova and B. Rosenhahn, "On calibration of a low-cost time-offlight camera," *Computer Vision - ECCV 2014 Workshops*, Springer International Publishing, pp. 415 - 427, 2015.
- [22] D. Ferstl, C. Reinbacher, G. Riegler, M. Rther, and H. Bischof, "Learning depth calibration of time-of-flight cameras," *Proceedings of the British Machine Vision Conference (BMVC)*, pp. 102.1 - 102.12, September, 2015.
- [23] A. Perez-Yus, E. Fernandez-Moral, G. Lopez-Nicolas, J. Guerrero, and P. Rives, "Extrinsic calibration of multiple RGB-D cameras from line observations," *IEEE Robot. Automat. Lett.*, vol. 3, no. 1, pp. 273 - 280, 2018.
- [24] N. Fukushima, "ICP with Depth Compensation for Calibration of Multiple ToF Sensors," *2018 - 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, Helsinki, Finland, pp. 1-4, 2018.
- [25] K. Desai, B. Prabhakaran and S. Raghuraman, "Skeleton-based Continuous Extrinsic Calibration of Multiple RGB-D Kinect Cameras," *Proceedings of the 9th ACM Multimedia Systems Conference*, pp. 250 - 257, June 2018.
- [26] K. Kim, B. Park, D. Kim, S. Kim, and Y. Seo, "Real-time 3D Volumetric Model Generation using Multiview RGB-D Camera," *Journal of Broadcast Engineering*, Vol. 25, No. 3, pp. 439-448, May 2020.

저 자 소 개



박 병 서

- 2019년 2월 : 광운대학교 경영학과 졸업(경영학사)
- 2019년 3월 ~ 현재 : 광운대학교 전자재료공학과(석박통합과정)
- ORCID : <https://orcid.org/0000-0003-3396-8264>
- 주관심분야 : 딥러닝, 2D/3D 영상 처리



강 지 원

- 2019년 2월 : 광운대학교 전자재료공학과 졸업(공학학사)
- 2019년 3월 ~ 현재 : 광운대학교 전자재료공학과(석사과정)
- ORCID : <https://orcid.org/0000-0002-8020-899X>
- 주관심분야 : 홀로그램, 딥러닝



이 솔

- 2018년 2월 ~ 현재 : 광운대학교 전자재료공학과 재학 중(학사과정)
- ORCID : <https://orcid.org/0000-0002-2154-1290>
- 주관심분야 : 3D 영상처리, 압축

저 자 소 개



박 정 탁

- 2016년 2월 ~ 현재 : 광운대학교 전자재료공학과 재학 중(학사과정)
- ORCID : <https://orcid.org/0000-0002-3825-2428>
- 주관심분야 : 3D 영상처리, 딥러닝



최 장 환

- 2016년 2월 ~ 현재 : 광운대학교 전자재료공학과 재학 중(학사과정)
- ORCID : <https://orcid.org/0000-0001-5808-5252>
- 주관심분야 : 3D 영상처리, 딥러닝, 디지털 홀로그램



김 동 욱

- 1983년 2월 : 한양대학교 전자공학과 졸업(공학사)
- 1985년 2월 : 한양대학교 공학석사
- 1991년 9월 : Georgia 공과대학 전기공학과(공학박사)
- 1992년 3월 ~ 현재 : 광운대학교 전자재료공학과 정교수
- ORCID : <https://orcid.org/0000-0001-6106-9894>
- 주관심분야 : 3D 영상처리, 디지털 홀로그램, 디지털 VLSI Testability, VLSI CAD, DSP설계, Wireless Communication



서 영 호

- 1999년 2월 : 광운대학교 전자재료공학과(공학사)
- 2001년 2월 : 광운대학교 일반대학원 졸업(공학석사)
- 2004년 8월 : 광운대학교 일반대학원 졸업(공학박사)
- 2005년 9월 ~ 2008년 2월 : 한성대학교 조교수
- 2008년 3월 ~ 현재 : 광운대학교 전자재료공학과 정교수
- ORCID : <https://orcid.org/0000-0003-1046-395X>
- 주관심분야 : 실감미디어, 2D/3D 영상 신호처리, 디지털 홀로그램, SoC설계