Journal of Korea Multimedia Society Vol. 22, No. 10, October 2019(pp. 1142-1148)
https://doi.org/10.9717/kmms.2019.22.10.1142

3차원 환경 복원을 위한 다중 키넥트의 마커리스 캘리브레이션

이 수 원†

Marker-less Calibration of Multiple Kinect Devices for 3D Environment Reconstruction

Suwon Lee†

ABSTRACT

Reconstruction of the three-dimensional (3D) environment is a key aspect of augmented reality and augmented virtuality, which utilize and incorporate a user's surroundings. Such reconstruction can be easily realized by employing a Kinect device. However, multiple Kinect devices are required for enhancing the reconstruction density and for spatial expansion. While employing multiple Kinect devices, they must be calibrated with respect to each other in advance, and a marker is often used for this purpose. However, a marker needs to be placed at each calibration, and the result of marker detection significantly affects the calibration accuracy. Therefore, a user-friendly, efficient, accurate, and marker-less method for calibrating multiple Kinect devices is proposed in this study. The proposed method includes a joint tracking algorithm for approximate calibration, and the obtained result is further refined by applying the iterative closest point algorithm. Experimental results indicate that the proposed method is a convenient alternative to conventional marker-based methods for calibrating multiple Kinect devices. Hence, the proposed method can be incorporated in various applications of augmented reality and augmented virtuality that require 3D environment reconstruction by employing multiple Kinect devices.

Key words: Kinect Calibration, Marker-less Calibration, 3D Environment Reconstruction, Augmented Reality, Augmented Virtuality

1. 서 론

2011년 마이크로소프트(MS)가 키넥트(Kinect)의 소프트웨어 개발 키트(software development kit, SDK)를 공개한 이래로 키넥트는 현재까지도 연구 및 산업분야에 다양한 형태로 활용되고 있다. 키넥트는 고화질의 색상(color) 영상과 가격에 비해 정교한 깊이(depth) 영상을 함께 획득할 수 있어서 키넥트를 이용하면 저렴한 가격으로 양질의 사용자 주변의 3차원 환경 복원(3D environment reconstruction)이 가능하다. 3차원 환경을 복원하는 것은 사용자의 주변 공간을 활용하는 증강현실(augmented reality, AR)[1,2]이나 증강가상(augmented virtuality, AV)[3,4]의 핵심기술이며, 키넥트를 활용한다면 손쉽게 구현가능하다. 뿐만 아니라 키넥트는 사용자 관절의 위치를 실시간으로 추적할 수 있는 응용 프로그램

※ Corresponding Author : Suwon Lee, Address: (52828) Jinjudae-ro 501, Jinju-si, Gyeongsangnam-do, Korea, TEL : +82-55-772-1394, FAX : +82-55-772-1389, E-mail : leesuwon@gnu.ac.kr
Receipt date : July 16, 2019, Revision date : Sep. 1, 2019
Approval date : Sep. 24, 2019

† Dept. of Computer Science and The Research Institute of Natural Science., Gyeongsang National University
※ This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (NRF-2018R1C1B5046098)

인터페이스(application programming interface, API)[5]를 제공한다는 장점으로 인해 인간과 컴퓨터와의 대화(human computer interaction, HCI)의 수단으로도 다양하게 활용되고 있다[6-8]. MS는 2014년에 키넥트의 두 번째 버전(version)을 출시하였으며 최근에 Azure 키넥트를 발표하였다.

키넥트는 한대만 활용해도 많은 응용이 가능하지만 여러 대의 키넥트를 이용하면 3차원 환경 복원의 밀도(density)를 높이거나 HCI의 정확도를 개선할 수 있다. 또한, 키넥트를 직렬로 혹은 마주보게 여러 대 배치하면 키넥트의 적용 범위를 확장하여 더 다양한 형태의 응용이 가능해진다.

다수의 키넥트를 활용할 때는 키넥트들이 사전에 서로 캘리브레이션(calibration)이 되어 있어야 하며, 이를 위해 마커(marker)를 이용하는 것이 전통적이다. 키넥트가 생성하는 3차원 영상에서 검출된 마커를 매개체로 하여 여러 대의 키넥트를 캘리브레이션 할 수 있다. 하지만 이 방법은 캘리브레이션을 할 때마다 마커의 출력물을 배치해야한다는 불편함을 가지고, 마커 검출의 결과가 캘리브레이션의 정확도에 상당한 영향을 끼친다.

본 논문에서는 마커를 필요로 하지 않는 새로운 방식의 다중 키넥트 캘리브레이션 방법을 제안한다. 제안한 방법은 마커가 필요하지 않아서 매우 편리하고, 쉽고, 빠르고, 정확한 캘리브레이션을 수행한다. 실제 구현을 통해 제안한 방법이 마커를 이용하는 캘리브레이션의 훨씬 편리한 대안이라는 것을 보인다. 본 연구는 다중 키넥트를 이용하여 3차원 환경 복원을 필요로 하는 증강현실이나 증강가상 등의 응용에 다양하게 활용될 것으로 기대된다.

본 논문의 구성은 다음과 같다. 2장에서 다중 키넥트 캘리브레이션의 전통적인 방법에 대해 소개한 다음, 3장에서 제안하는 방법에 대해 설명한다. 여기에서는 제안한 방법의 전체적인 개요와 개략적인 캘리브레이션, 정확도 개선 및 3차원 환경 복원 순으로 서술한다. 그리고 4장에서는 구현 방법과 결과에 대해 서술하면서 제안한 방법의 효율성을 보이고 5장에서 결론을 맺는다.

2. 전통적인 방법

다중 키넥트 캘리브레이션의 전통적인 방법은 이어지는 2.1장과 2.2장에서 설명할 두 절차를 순차적으로 수행하여 이루어진다. 먼저 각각의 키넥트가 3차원 영상을 생성하게 만든 후, 모든 키넥트의 겹치는 시야에 배치한 마커의 위치가 원점이 되도록 3차원 영상의 모든 좌표를 변환한다.

2.1 키넥트의 3차원 영상 생성

다중 키넥트 캘리브레이션은 각각의 키넥트가 생성하는 3차원 영상을 기반으로 이루어진다. 따라서 다중 키넥트 캘리브레이션을 하기 위해서는 사전에 각각의 키넥트가 독립적으로 3차원 영상을 생성하고 있어야한다.

키넥트는 색상 영상과 깊이 영상의 두 채널(channel)로 구성되며, 3차원 영상은 이 두 채널의 캘리브레이션을 통해 생성된다. 이는 두 번의 매핑(mapping)을 통해 이루어진다. 먼저 깊이 영상에서 카메라 공간으로의 매핑을 통해 깊이 영상의 매 픽셀마다 카메라 공간으로 변환되었을 때의 3차원 좌표를 계산한다. 그 후, 깊이 영상에서 색상 영상으로의 매핑을 통해 깊이 영상의 매 픽셀마다 해당하는 색상 값을 대입한다. 모든 3차원 좌표를 대입된 색상 값을 이용해 3차원 공간으로 출력하여 3차원 영상을 생성한다.

2.2 마커를 이용한 다중 키넥트 캘리브레이션

2.1장에서 생성된 3차원 영상은 키넥트의 위치를 원점으로 하는 카메라 공간상의 3차원 영상이다. 따라서 다중 키넥트를 캘리브레이션을 하기 위해서는 각각의 3차원 영상 사이의 변환을 수행해야한다. 이를 위해서는 모든 3차원 영상에서 동일하게 존재하는 좌표들을 찾아야하며, 전통적으로는 마커를 이용한다. 일반적으로 마커는 사각형이지만 오각형 이상의 마커를 이용하여 캘리브레이션의 정확도를 높일 수 있다[9].

키넥트의 시야가 겹치는 위치에 마커의 출력물을 배치하고, 마커 검출 알고리즘[10]을 이용하여 색상 영상에서의 마커를 검출한다. 먼저 모든 윤곽(contour)을 찾아낸 후, 윤곽의 볼록함(convexity), 꼭지점(vertex)의 개수와 꼭지점이 이루는 넓이를 기준으로 걸러서 남은 윤곽들을 마커의 후보 영역으로 결정한다. 그 후, 마커 내부의 양식(pattern)을 평가해 최종 마커로 검출한다. 검출된 마커의 꼭지점 좌표에 해당하는 3차원 영상에서의 좌표는 카메라 공

간에서의 마커의 위치를 의미한다. 3차원 영상에서의 꼭지점 좌표들을 마커의 위치가 원점이 되도록 대입하고, 직각 프로크루스테스(orthogonal Procrustes) 알고리즘[11]을 이용해 변환 행렬(transformation matrix)을 계산한다. 계산된 행렬은 키넥트의 외부 파라메터(extrinsic parameter)로서 키넥트의 위치를 마커의 위치로 변환하는 3차원 회전(rotation)과 3차원 평행이동(translation)을 포함한다. 3차원 영상의 모든 좌표에 자신이 속한 키넥트의 외부 파라메터를 곱하여 좌표를 변환한다. 이 과정에서 다수의 키넥트가 캘리브레이션이 되며, 모든 변환된 좌표를 동일한 3차원 공간으로 출력하여 최종 3차원 환경을 복원한다.

3. 제안한 방법

3.1 제안한 방법의 개요

본 논문에서는 새로운 방식의 다중 키넥트 캘리브레이션 방법을 제안한다. 제안한 방법은 마커를 이용하는 전통적인 방법에서 벗어나 별도의 마커를 필요로 하지 않으면서 사용하기 쉽고, 편리하게 빠른 속도로 정확한 캘리브레이션을 수행한다. Fig. 1은 제안한 방법의 전체적인 흐름도를 나타낸다.

먼저, 2.1장에서 소개한 방법을 이용하여 각각의 키넥트가 3차원 영상을 생성하도록 한다. 그 후, 각각의 키넥트에서 사용자의 관절을 추적한다. 관절의 위치를 기반으로 특이값 분해(singular value decomposition, SVD) 알고리즘을 이용하여 키넥트와 키넥트 사이의 외부 파라메터인 변환 행렬을 계산한다. 정확도를 향상시키기 위해 변환된 3차원 영상들을 입력으로 하여 반복적인 최근점(iterative closest points, ICP) 알고리즘으로 결과를 정제(refinement)

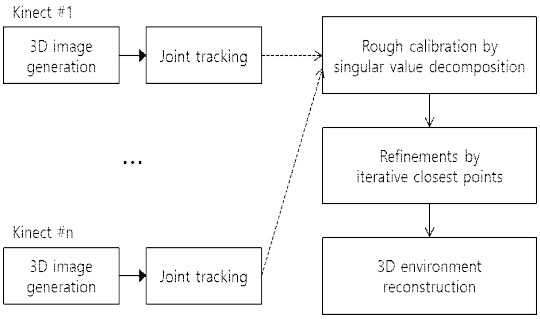


Fig. 1. Flowchart for the proposed method.

한다. 이어지는 3.2장과 3.3장에서는 제안한 방법을 자세하게 설명한다.

3.2 관절 추적 알고리즘을 이용한 다중 키넥트의 개략적 캘리브레이션

제안한 방법은 마커 대신에 사용자의 관절 위치를 이용한다. 키넥트의 시야가 겹치는 공간에 사용자를 위치시키고 동일한 프레임(frame) 상에서 사용자의 관절을 추적한다. 키넥트는 무작위 트리(randomized trees) 알고리즘[12]을 이용하여 깊이 영상에서 사용자의 관절을 추적한다[13,14].

깊이 영상에서 추적된 관절의 위치는 2.1장에서 소개한 매핑을 통해 3차원 영상에서의 관절의 좌표로 변환한다. 각각의 키넥트가 생성하는 3차원 영상은 고유한 원점을 가지기 때문에 동일한 관절의 위치일지라도 모두 다른 좌표를 가진다. 점들의 집합 사이의 변환 관계를 구하기 위해 먼저, 점들의 중점을 아래와 같이 계산한다.

$$C = \frac{1}{N}\sum_{i=1}^{N} P^i \tag{1}$$

P는 3차원 공간상에서의 점을 표현하기 위한 3차원 벡터(vector)이며, N은 점들의 개수이다. 중점을 이용하여 점들의 두 집합을 하나의 원점으로 맞춰주고 공분산 행렬(covariance matrix)을 아래와 같이 계산한다.

$$H = \sum_{i=1}^{N} (P_A^i - C_A)(P_B^i - C_B)^T \tag{2}$$

A와 B는 서로 다른 키넥트를 구별한다. SVD 알고리즘을 이용하여 아래와 같이 3차원 회전 행렬을 도출한다.

$$[U, S, V] = SVD(H) \tag{3}$$

$$R = VU^T \tag{4}$$

마지막으로 3차원 평행이동 벡터는 아래와 같이 계산한다.

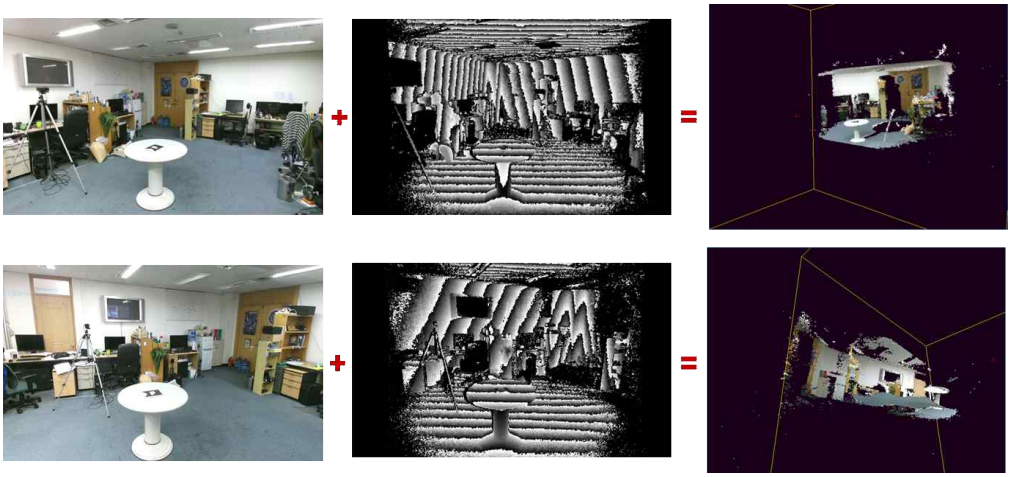
$$T = -R \times C_A + C_B \tag{5}$$

계산된 회전 행렬과 평행이동 벡터는 키넥트의 외부 파라메터를 구성한다.

3.3 ICP 알고리즘을 이용한 정확도 개선 및 3차원 환경 복원

캘리브레이션의 정확도는 관절 추적 알고리즘의

Fig. 2. Experimental environment to verify the proposed method.



(a) (b) (c)

Fig. 3. 3D image generation of two Kinects without calibration. (a) color image, (b) depth image and (c) 3D image.

성능에 의해 결정된다. 전통적인 방법에서 마커 검출 알고리즘이 캘리브레이션의 성능을 결정하는 것과 같은 이치이다. 캘리브레이션의 정확도를 개선하기 위해 ICP 알고리즘[15]을 이용한다. ICP 알고리즘은 개략적으로 정렬된 두 포인트 집합이 있을 때, 서로 간에 가장 가까운 포인트 간 거리의 합이 최소가 되도록 반복적으로 포인트들을 일괄적으로 회전 및 이동시키는 알고리즘이다. 이는 각각의 포인트마다 반대 집합에 있는 포인트들 중에서 가장 가까운 포인트를 찾는 과정과 이들의 거리를 줄일 수 있도록 일괄적으로 변환해주는 행렬을 찾는 두 과정의 반복으로 이루어진다.

개략적으로 정렬된 포인트 집합을 준비하기 위해 이전 과정에서 계산한 외부 파라메터를 이용한다. 이는 키넥트와 키넥트 사이의 관계이기 때문에 메인(main)으로 사용할 키넥트를 지정해야한다. 메인 키넥트를 제외한 나머지 서브(sub) 키넥트의 3차원 영상의 모든 좌표에 메인 키넥트로 변환되는 외부 파라메터를 곱하여 좌표를 변환한다. 서브 키넥트와 메인 키넥트가 생성하는 3차원 영상을 포인트 클라우드로 두고 ICP 알고리즘을 수행하여 외부 파라메터를 정제한다. 이 과정에서 캘리브레이션의 정확도가 개선되고, 서브 키넥트의 위치에서 메인 키넥트의 위치로 변환되는 최종 외부 파라메터가 도출된다.

서브 키넥트의 3차원 영상의 모든 좌표에 최종 외부 파라메터를 곱하여 좌표를 변환하고, 모든 변환된 좌표를 동일한 3차원 공간으로 출력하여 최종 3차원 환경을 복원한다.

4. 구현 및 결과

두 대의 키넥트를 이용하여 제안한 방법을 구현하였다. Fig. 2와 같이 두 대의 키넥트를 겹치는 부분이 있도록 배치하고, 각각의 노트북에 연결하였다. 키넥트의 영상을 TCP 소켓을 이용하여 공유기를 거쳐 PC로 스트리밍(streaming)을 하는 환경을 구축하였다.

Fig. 4. 3D environment reconstruction of two Kinects with calibration.

Table 1. Comparison of calibration error

Test Set	Method	
	Conventional method	Proposed method
	Error(cm)	Error(cm)
1	3.32	3.26
2	12.78	9.57
3	4.47	4.10
4	6.02	5.43
5	2.24	1.88
Average	5.76	4.84

키넥트 API의 ICoordinateMapper 인터페이스(interface)를 이용하여 깊이 영상과 색상 영상을 매핑하는 과정을 구현하여 키넥트가 3차원 영상을 생성하도록 하였다. Fig. 3과 같이 깊이 영상과 색상 영상이 결합하여 3차원 영상을 생성하지만 키넥트는 서로 캘리브레이션이 되지 않은 상태이다.

두 키넥트의 시야가 겹치는 공간에 사용자를 위치시키고 키넥트 API의 IBodyFrame 인터페이스를 이용하여 사용자의 관절을 추적하고 각 관절의 위치를 3차원 영상에서의 좌표로 변환하였다. Fig. 4는 3차원 영상에서의 관절의 위치에 대한 좌표를 이용해 만든 사용자의 뼈대를 하나의 3차원 공간에 투영한 결과를 보여준다. 현재는 두 키넥트가 서로 캘리브레이션이 되지 않아 사용자의 뼈대가 두개로 나타남을 볼 수 있다.

사용자 관절의 위치를 기반으로 SVD 알고리즘을 구현하여 개략적인 캘리브레이션을 수행하였다. 실제 구현에서는 오류를 최소화하기 위해서 약 1초 동안의 프레임을 누적하였고, 각각의 프레임에서 계산한 값들의 평균을 최종 결과로 사용하였다. 두 대의 키넥트 중 한대를 메인으로 지정하고, 나머지 한대의 모든 좌표를 캘리브레이션 결과를 이용하여 메인 키넥트의 좌표로 변환하였다. 이후, ICP 알고리즘을 수행하여 이를 정제하였으며 10회 반복하는 것으로 구현하였다. 기본 ICP 알고리즘의 구현은 [16]를 참고하였다. Fig. 5는 제안한 방법에 의해 캘리브레이션된 두 키넥트로 복원된 3차원 환경을 보여준다. 마커를 이용한 전통적인 방법과 비교하기 위해 메인 키넥트의 원점을 마커의 위치로 변환하였다.

캘리브레이션의 정확도를 측정하기 위해 가로와 세로가 각각 10개의 격자로 이루어진 체크 보드를 이용하였다. 키넥트의 깊이 영상에서 체크 보드의 모서리를 검출하고, 해당 위치에 매핑된 3차원 좌표를 추출하였다. 이를 캘리브레이션이 된 두 키넥트에서 각각 수행한 후, 두 키넥트 사이의 동일한 모서리의 3차원 좌표의 오차를 cm단위로 변환하였다. 체크 보

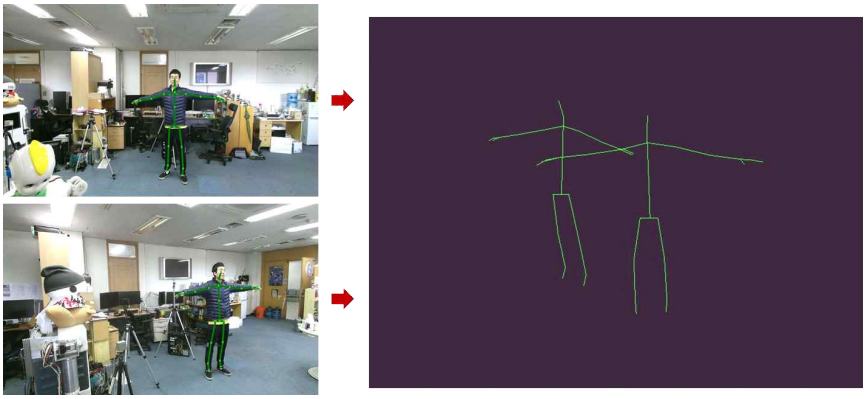


Fig. 5. 3D image generation of two Kinects without calibration.

(a) (b)

Fig. 6. Comparison of accuracy of 3D environment reconstruction. (a) conventional method and (b) proposed method.

드의 위치를 달리하며 총 5번의 테스트를 수행하였고, 각 테스트 마다 총 100개의 모서리에 대한 평균 오차를 계산하였다. 표 1은 제안한 방법과 전통적인 방법을 비교한 결과이다. 모든 테스트에서 제안한 방법이 전통적인 방법에 비해 정확하게 캘리브레이션이 되었음을 보여준다. Fig. 6은 제안한 방법이 전통적인 방법에 비해 정확도 면에서 우수함을 시각적으로 보여준다. ICP 알고리즘에 의해 캘리브레이션의 정확도가 향상되면서 동일한 물체가 마치 두개로 보이던 현상이 많이 줄어들었음을 볼 수 있다.

5. 결 론

본 논문에서는 마커를 필요로 하지 않으면서 사용하기 쉽고, 빠르고, 정확한 다중 키넥트 캘리브레이션 방법을 제안하였다. 사용자 관절의 위치를 기반으로 SVD 알고리즘을 이용하여 키넥트와 키넥트 사이의 개략적인 캘리브레이션을 수행한 후, 그 결과를 ICP 알고리즘을 이용하여 정제하여 정확도를 개선하였다. 실제 구현을 통해 제안한 방법이 다수의 키넥트를 캘리브레이션하기 위한 방법으로 사용될 수 있음을 검증하였고, 마커를 이용하는 전통적인 방법보다 훨씬 편리한 대안이라는 것을 보였다. 뿐만 아니라 ICP 알고리즘에 의해 정확도면에서도 향상된 점을 확인할 수 있었다. 본 연구는 다중 키넥트를 이용하여 3차원 환경 복원을 필요로 하는 증강현실이나 증강가상 등의 응용에 다양하게 활용될 것으로 기대된다.

REFERENCE

[1] Y. Zhu, K. Zhu, Q. Fu, X. Chen, H. Gong, J. Yu, et al., "SAVE: Shared Augmented Virtual Environment for Real-Time Mixed Reality Applications," *Proceedings of ACM SIGGRAPH Conference on Virtual-Reality Continuum and Its Applications in Industry*, pp. 13-21, 2016.

[2] E. Bostanci, N. Kanwal, and A.F. Clark, "Augmented Reality Applications for Cultural Heritage Using Kinect," *Human-Centric Computing and Information Sciences*, Vol. 5, No. 1, pp. 1-18, 2015.

[3] D. Nahon, G. Subileau, and B. Capel, ""Never Blind VR" Enhancing the Virtual Reality Headset Experience with Augmented Virtuality," *Proceedings of IEEE Virtual Reality*, pp. 347-348, 2015.

[4] P.Q. David, N.P. Barteld, and F.K. Brian, "Augmented Auralization: Complementing Auralizations with Immersive Virtual Reality Technologies," *Proceedings of International Symposium on Music and Room Acoustics*, pp. 1-10, 2016.

[5] Reference Documentation of the Kinect for Windows SDK 2.0, http://docs.microsoft.com/en-us/previous-versions/windows/kinect/ (accessed July 19, 2019).

[6] N. Nikolakis, V. Maratos, and S. Makris, "A Cyber Physical System (CPS) Approach for Safe Human-Robot Collaboration in a Shared Workplace," *Robotics and Computer-Integrated Manufacturing*, Vol. 56, No. 1, pp. 233-243, 2019.

[7] B. Muller, W. Ilg, M.A. Giese, and N. Ludolphs, "Validation of Enhanced Kinect Sensor Based Motion Capturing for Gait Assessment," *PloS one*, Vol. 12, No. 4, pp. 1-18, 2017.

[8] Y. Choi, J. Tang, S. Jang, and S. Kim, "User Customizable Hit Action Recognition Method using Kinect," *Journal of Korea Multimedia Society*, Vol. 18, No. 4, pp. 557-564, 2015.

[9] T. Rybus, J. Nicolau-Kuklinski, K. Seweryn, T. Barcinski, M. Ciesielska, K. Grassmann, et al., "New Planar Air-bearing Microgravity

Simulator for Verification of Space Robotics Numerical Simulations and Control Algorithms," *Proceedings of Symposium on Advanced Space Technologies in Robotics and Automation*, pp. 1-8, 2013.

[10] X. Zhang, S. Fonz, and N. Navab, "Visual Marker Detection and Decoding in AR Systems: A Comparative Study," *Proceeding of International Symposium on Mixed and Augmented Reality*, pp. 97-106, 2002.

[11] P. Schnemann, "A Generalized Solution of the Orthogonal Procrustes Problem," *Psychometrika*, Vol. 31, No. 1, pp. 1-10, 1966.

[12] V. Lepetit, P. Lagger, and P. Fua, "Randomized Trees for Real-Time Keypoint Recognition," *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 775-781, 2005.

[13] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, et al., "Real-Time Human Pose Recognition in Parts from Single Depth Images," *Communications of the ACM*, Vol. 56, No. 1, pp. 116-124, 2013.

[14] J. Shotton, A. Fitzgibbon, A. Blake, A. Kipman, M. Finocchio, M. Moore, et al., "Real-Time Human Pose Recognition in Parts from Single Depth Images," *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1297-1304, 2011.

[15] P.J. Besl and N.D. McKay, "Method for Registration of 3-D Shapes," *Sensor Fusion IV: Control Paradigms and Data Structures*, Vol. 1611, No. 1, pp. 586-606, 1992.

[16] M. Kowalski, J. Naruniec, and M. Daniluk, "LiveScan3D: A Fast and Inexpensive 3D Data Acquisition System for Multiple Kinect v2 Sensors," *Proceeding of IEEE International Conference on 3D Vision*, pp. 318-325, 2015.

이 수 원

2012년 한국과학기술원(공학석사)
2017년 한국과학기술원(공학박사)
2018년~현재: 경상대학교 컴퓨터과학과 조교수
관심분야 : 증강현실(Augmented Reality), 컴퓨터비전(Computer Vision)