

포즈 추적 센서를 활용한 모바일 증강 가시화 기술

이동춘, 김항기, 이기석

한국전자통신연구원

{bluepine, hangkee, mvr_lks}@etri.re.kr

Mobile Augmented Visualization Technology Using Vive Tracker

Dong-Chun Lee, Hang-Kee Kim, Ki-Suk Lee

ETRI(Electronics and Telecommunications Research Institute)

요 약

본 논문은 2개의 포즈(위치 및 회전) 추적 센서를 사용하여 마네킹 모델위에 3차원의 가상 인체를 증강하는 모바일 증강 가시화 기술에 대해서 소개한다. 증강 가시화를 위해 사용된 종래의 카메라 트래킹 기술은 카메라 영상을 사용하기 때문에 카메라의 떨림이나 빠른 이동시 카메라의 포즈 계산에 실패하는 단점이 있으나, 바이브 트래커를 이용하게 되면 이러한 단점을 극복할 수 있다. 또한 증강하고자 하는 객체인 마네킹의 위치가 바뀌거나 회전을 하게 되더라도 마네킹에 부착된 포즈 추적 센서를 사용하여 증강 가시화가 가능한 장점이 있으며 무엇보다 카메라 트래킹을 위한 부하가 없다는 장점을 가진다.

ABSTRACT

This paper introduces a mobile augmented visualization technology that augments a three-dimensional virtual human body on a mannequin model using two pose(position and rotation) tracking sensors. The conventional camera tracking technology used for augmented visualization has the disadvantage of failing to calculate the camera pose when the camera shakes or moves quickly because it uses the camera image, but using a pose tracking sensor can overcome this disadvantage. Also, even if the position of the mannequin is changed or rotated, augmented visualization is possible using the data of the pose tracking sensor attached to the mannequin, and above all there is no load for camera tracking.

Keywords : Vive tracker(바이브 트래커), Augmented reality(증강현실), Mobile augmented reality(모바일 증강현실), 이기종 좌표계 통합

Received: Jul. 08. 2021 Revised: Oct. 01. 2021

Accepted: Oct. 06. 2021

Corresponding Author: Dong-Chun Lee(ETRI)

E-mail: bluepine@etri.re.kr

ISSN: 1598-4540 / eISSN: 2287-8211

© The Korea Game Society. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서 론

증강현실은 카메라로부터 들어오는 실사의 영상과 컴퓨터를 통해 만들어진 가상의 영상을 정합하여 사용자에게 부가의 정보를 제공하는 기술이다. 이때 실제 객체와 가상의 영상에서 만들어진 객체가 얼마나 오차 없이 정합 되는지는 증강현실을 평가하는 중요 요소 중 하나이다. 증강현실에서 정합되는 두 영상의 오차를 줄이기 위해서는 실시간 변화하는 카메라 포즈(위치 및 회전)에 대한 정확한 추적이 필요하다.

증강현실에서 사용하는 카메라 추적 기술은 초기엔 카메라 영상에서 마크를 인식하고, 카메라의 움직임에 따라 변화하는 마크의 모양을 기반으로 카메라 포즈를 계산하는 방식을 사용하였으나, 2008년 공간인식 기반 트래킹 기술인 SLAM (Simultaneous Localization and Mapping)을 사용한 PTAM과 ORB-SLAM이 소개되면서 비마커 기반 카메라 추적 기술을 많이 사용하게 되었다 [1,2,8]. 공간인식 기반 트래킹 기술 초기에 단일 카메라를 사용하였으나[2,7,16,17] 카메라 트래킹의 정확도를 높이기 위해 스테레오 카메라를 사용하였다[8,9,18]. 이후 카메라 영상뿐 아니라, 스마트폰에 내장된 IMU 센서 데이터를 활용하여 카메라의 포즈를 트래킹하는 VIO(Visual Inertial Odometry) 기술을 사용한 모바일 트래킹 기술이 2017년 출시되었다[3,4,5,6]. 2017년 애플은 IOS에서 구동되는 VIO 기반 증강현실 엔진인 ARKit을 출시하였고, 2018년 구글은 안드로이드에서 동작하는 증강현실 엔진 ARCore를 출시하였다[10,11]. 최근에는 증강현실에 인공지능 기술을 접목하여 실사 영상에 대한 깊이맵을 자동으로 예측하는 기술이 소개되고 있다[12,13].

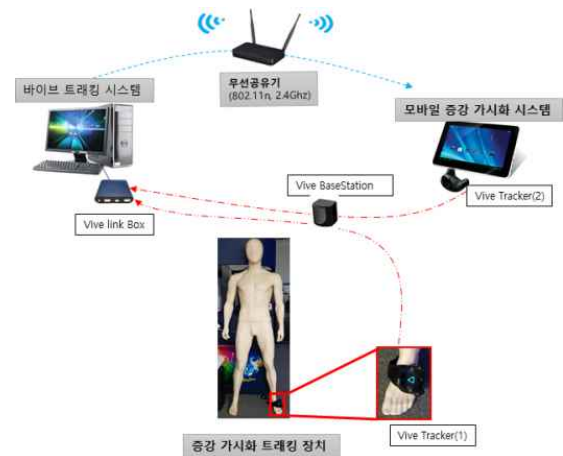
그렇지만 앞서 언급한 대부분의 카메라 트래킹 기술은 카메라 영상을 사용하여 카메라의 위치를 추적하기 때문에 카메라의 움직임이 빠르게 되면 트래킹이 안되는 단점이 있다. 또한 증강하고자 하

는 대상 물체의 변화가 적고, 고정된 실사 객체를 대상으로 적용하기는 적합하나, 움직이는 실사 객체의 위치를 트래킹하여 증강영상을 생성하기엔 힘들다는 단점이 있다. 본 논문에서는 이동이 가능한 마네킹 객체에 가상의 3차원 인체 정보를 증강하기 위해서 포즈 추적 센서 장치인 바이브 트래커를 활용한 모바일 증강 가상화 기술에 대해서 소개한다.

2. 본 론

2.1 포즈 추적 센서 활용 모바일 트래킹

시스템 구성



[Fig. 1] Mobile Tracking System Configuration

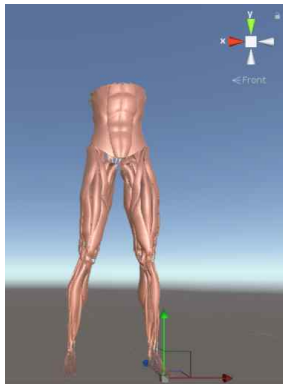
[Fig. 1]은 모바일 트래킹 시스템 구성을 나타낸다. 모바일 트래킹 시스템은 바이브 트래킹 시스템과 모바일 증강 가상화 시스템 2개로 구성되어 있다. 바이브 트래킹 시스템은 발에 부착된 바이브 트래커와 모바일 단말에 부착된 바이브 트래커의 포즈를 추적하여 모바일 증강 가상화 시스템에 전달한다. 모바일 증강 가상화 시스템은 바이브 트래킹 시스템에서 전달받은 트래커의 포즈 데이터를 사용하여 증강 영상을 생성한다. 모바일 증강 가상화 시스템에서는 모바일 단말의 포즈 추적을 위해서 모바일 단말에 바이브 트래커를 부착하여 사용

한다. [Fig. 2]는 모바일 증강 가시화 시스템의 하드웨어 구성을 보여준다.



[Fig. 2] Mobile augmented visualization system hardware configuration

[Fig. 3]는 마네킹 발에 부착된 바이브 트래커의 포즈 데이터를 반영하여 실사 영상에 증강 가시화되는 3차원 가상 인체 모델 데이터를 보여준다. 마네킹 발에 부착된 바이브 트래커의 움직임과 동기화를 위해 3차원 마네킹 모델의 좌표축 중심을 마네킹의 왼쪽 발바닥으로 설정하였다.



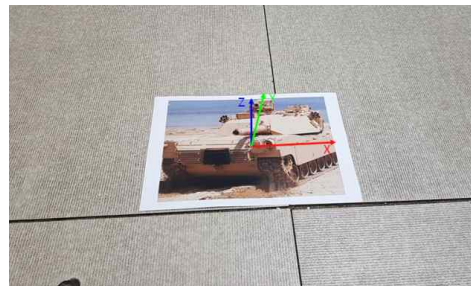
[Fig. 3] 3D virtual human body model in Unity3D

2.2 시스템 좌표계

본 논문에서는 두 개의 좌표계 시스템(바이브 트래커 좌표계, 모바일 디바이스 좌표계)을 가지고 있다. 본 논문에서는 이 두 좌표계를 최종적으로 바이브 좌표계로 통합하여 사용하는데, 통합을 위해서는 모바일 단말에 부착된 바이브 트래커로부터 모바일 단말에 부착된 카메라 렌즈까지의 상대 포즈가 계산되어야 한다. 바이브 트래커의 좌표계는 왼손 좌표계를 사용하였고, 실세계의 중력 방향의

반대 방향을 Y축으로 사용하였다. 이 때문에 지면 바닥은 XZ평면으로 구성되어진다.

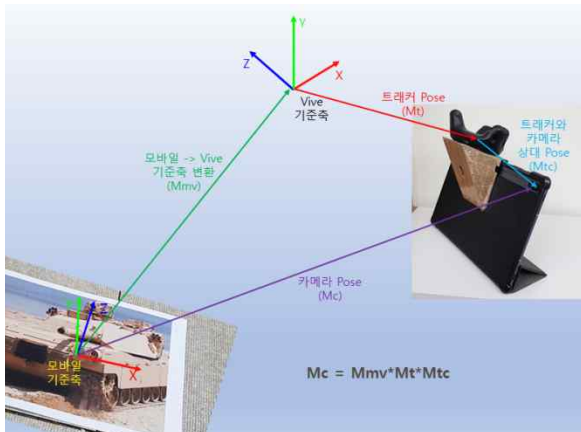
모바일 디바이스 좌표계는 시스템 좌표계 통합을 위해 일시적으로 사용하는 좌표계로 모바일 디바이스에 부착된 바이브 트래커가 모바일 디바이스에 부착된 카메라로부터 상대적으로 얼마나 떨어져 있고, 얼마나 회전되어 있는지를 측정하기 위해서 사용되어진다. 본 논문에서는 모바일 디바이스 좌표계를 위해서 뷰포리아의 마커 인식 기술을 사용하였다[14]. 이 때문에 마커가 놓여진 지면 바닥 공간이 XZ평면을 이루고, 지면의 수직인 실세계 중력 방향의 반대 방향이 Y축으로 구성되는 왼손 좌표계가 사용된다. [Fig. 4]은 탱크 이미지 마커를 사용한 모바일 좌표계를 보여준다.



[Fig. 4] Marker based mobile coordinate

2.3 시스템 좌표계 통합

모바일 단말에 부착된 바이브 트래커를 사용하여 증강 가시화를 하기 위해서는 모바일 단말에 부착된 바이브 트래커와 카메라의 상대 포즈(Mtc)를 알아야 한다. 카메라와 바이브 트래커의 상대 포즈(Mtc) 계산식은 아래와 같다.



[Fig. 5] Two coordinate system transformation process

$$Mc = Mmv * Mt * Mtc \text{ ----- (1)}$$

Mc: 모바일 디바이스 좌표계(모바일 기준축)상에서의 카메라의 포즈값

Mmv: 모바일 디바이스 좌표계를 바이브 트래커 좌표계로 변환하는 변환 매트릭스

Mt: 바이브 트래커 좌표계(Vive 기준축)상에서의 모바일 디바이스에 부착된 Vive 트래커 포즈값

Mtc: 바이브 트래커와 카메라의 상대 포즈값

[Fig. 5]는 두 좌표계의 변환 과정을 보여주는 그림이다. 수식 (1)은 아래의 수식으로 표시될 수 있다.

$$Mc - Mmv * Mt * Mtc = 0 \text{ ----- (2)}$$

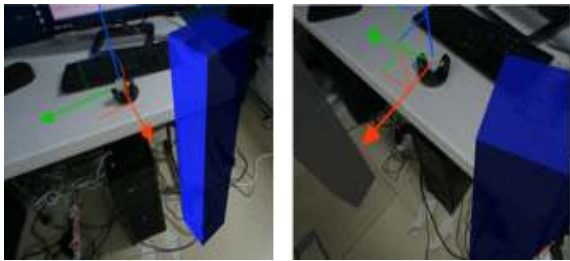
수식 (2)에서 Mc는 퓨포리아 마커 인식 시스템을 통해 계산되어지는 값이고, Mt는 바이브 시스템에 의해 측정되어지는 값으로 두 값은 측정을 통해 알 수 있는 값이다. 모바일 좌표계에서 바이브 좌표계로 변환하는 Mmv와 바이브 트래커와 모바일 카메라의 상대 포즈값 Mtc는 회전과 위치에 대한 6개의 변수 값을 가진다. 따라서 위의 수식(2)는 총 12개의 변수 값을 가지게 된다. 그렇지만 본 시스템에서 사용하는 두 좌표계는 모두 지면 바닥을 XZ평면으로 하기 때문에 두 좌표계의 변환 매트릭스 값인 Mmv는 Y축에 대한 회전 값

만 고려하면 되고, 이 때문에 Mmv는 4개의 변수 값을 가지게 된다. 결과적으로 위의 수식 (2)는 총 10개의 변수로 구성되어진다.

위의 수식 (2)를 풀기 위해 모바일 단말기의 위치 및 방향을 달리하면서 7~10개의 데이터값(바이브 트래커 포즈값, 퓨포리아 카메라 포즈값)을 측정하였다. 측정된 값들을 이용하여 위의 수식 (2)의 좌변의 절대치를 최소화하는 해를 구함으로써 트래커와 카메라의 상대 포즈 매트릭스 값을 계산하였다. 본 논문에서는 트래커와 카메라의 상대 포즈 매트릭스 값을 구할 때 회전값을 먼저 계산하고, 이후 위치 값을 계산하였다. 회전과 위치를 구분하여 따로 계산하게 되면 한번 계산시 사용되는 변수의 개수가 최대 6개를 넘지 않게 되어 더 효율적인 계산이 가능하다.

3. 실험 및 결과

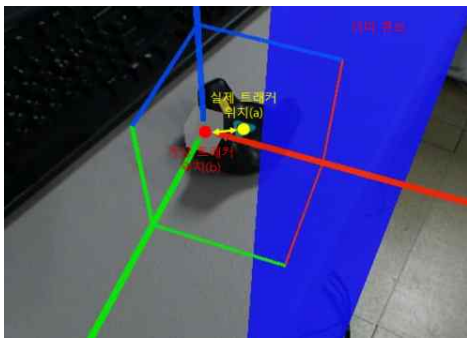
본 논문에서는 마네킹과 3차원 인체 모델을 정합하기 위한 포즈 추적 센서로 바이브 트래커를 사용하였다. 모바일 디바이스에 부착된 바이브 트래커와 모바일 디바이스에 부착된 카메라 렌즈와의 상대 포즈를 계산하는 과정이 필요하나, 이 과정은 바이브 트래커가 고정되어 있다면 한번만 계산하면 되고 이후 바이브 시스템이나 모바일 시스템이 초기화되거나 움직이더라도 다시 계산할 필요가 없었다. 또한 마네킹의 위치를 옮기거나 회전을 시키더라도 추가 작업 없이 실사의 마네킹과 가상의 인체 모델을 정합할 수 있었다. 본 논문에서는 갤럭시 시 탭 S6 기종의 모바일 디바이스를 사용하여 실험하였고, 바이브 트래킹 시스템은 Vive Pro를 사용하여 실험하였다[15].



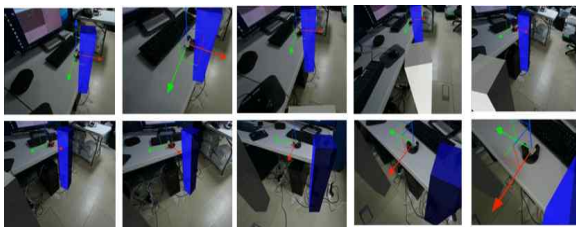
[Fig. 6] Augmentation result of 3D coordinate model and vive tracker

[Fig. 6]은 RGB 화살표로 표시되는 가상의 3차원 객체와 실세계의 바이브 트래커를 정합한 결과 화면으로 통합된 좌표계의 정확도를 알 수 있다. 본 논문에서는 정합의 정확도를 정확히 계산하기 위해서 약 30초 동안 카메라의 위치 및 방향을 이동하며 캡처된 증강 영상에 대하여 실제 바이브 트래커의 위치와 가상의 바이브 트래커 위치의 차이를 계산하였다.

[Fig. 7]은 정합 오차를 계산하는 방법을 설명하는 그림으로 정합 오차는 실제 바이브 트래커 위치(a)와 가상의 바이브 트래커 위치(b)의 차이로 정의될 수 있다. [Fig. 8]는 정합오차 계산에 사용된 이미지 시퀀스 중 일부를 보여준다.



[Fig. 7] Registration error definition



[Fig. 8] Images for calculating registration error

VGA 해상도 기준 실제 바이브 트래커와 가상의 바이브 트래커 정합 오차(픽셀 단위) 결과는 [Table 1]과 같다.

[Table 1] Registration error in pixels

frame	실제 트래커 위치(a)		가상 트래커 위치(b)		정합 오차	
	u	v	u	v	du	dv
30	335	166	323	168	12	2
60	284	141	276	141	8	0
90	299	143	284	147	15	4
120	316	136	295	139	21	3
150	311	136	295	139	16	3
180	291	98	276	98	15	0
210	270	95	260	94	10	1
240	266	68	255	66	11	2
...
1020	289	136	273	138	16	2

total du = 339, total dv = 160

average du = 9.97, average dv=4.7

[Fig. 9]는 마네킹의 발에 Vive 트래커를 부착한 후 마네킹과 3차원 인체 근육 모델의 정합한 결과 화면이다. 3차원 인체 모델은 마네킹의 정면과 측면 두 개의 영상을 바탕으로 표준 근육모형을 변형해서 마네킹의 체형에 맞게 제작했다.



[Fig. 9] Augmentation result of mannequin and 3D human muscle model

4. 결 론

본 논문에서는 2개의 포즈 추적 센서(바이브 트

래커)를 활용하여 실사의 마네킹에 가상의 3차원 마네킹 영상을 정합하는 방법에 대해서 제시하였다. 기존에 사용하고 있는 카메라 영상을 사용한 증강현실의 경우 카메라로부터 들어오는 영상으로부터 ORB 혹은 Surf 등의 특징점을 추출하고, 추출된 특징점의 매칭을 통해 매칭되어진 특징점을 이용해서 카메라 포즈를 계산하는데 많은양의 이미지 처리 연산이 필요한 단점이 있었다. 그렇지만 본 논문에서 제안하는 방법은 바이브 트래킹 하드웨어 시스템에서 자동으로 계산되는 바이브 트래커의 포즈(Mt)와 시스템 진처리 과정에서 구해진 바이브 트래커와 카메라 렌즈의 상대 포즈값(Mtc)을 곱하여 모바일 단말의 카메라 포즈를 계산할 수 있어 카메라 트래킹을 위한 연산량은 거의 없다고 할 수 있다. 실제로 종전의 ORB-SLAM의 경우 Tracking 단계를 수행하는데 평균 30.57ms의 수행 시간을 보여주고 있다[2].

본 논문에서 제시한 바이브 트래커를 사용하여 모바일 증강 가상화 기술은 정합하고자 하는 객체의 위치가 이동하거나 회전이 되어도 객체에 부착된 바이브 트래커 센서 정보를 활용하여 쉽게 정합이 가능하다. 또한 카메라 영상을 사용하여 카메라 트래킹을 하지 않아 조명의 밝고 어두움에 영향이 없고 특히 카메라의 빠른 움직임과 흔들림으로 인한 카메라 영상의 선명도가 떨어지더라도 정합이 가능한 장점을 가진다. 그렇지만 바이브 트래커가 트래커를 추적하는 센서로부터 가리워지면 추적이 불가능하고, 바이브 트래커 센싱이 가능한 공간이 최대 5m x 5m로 제한된 공간에서 트래킹이 가능하다는 한계는 존재한다. 본 논문에서는 포즈 추적 센서 장치로 바이브 트래커를 사용하였으나 본 논문에서 제시한 기술을 사용한다면 동작 캡처용으로 많이 사용하는 OptiTrack 장치를 활용한 모바일 증강 가상화 시스템 구성이 가능하다.

ACKNOWLEDGMENTS

이 논문은 2021년도 정부(과학기술정보통신부)

의 재원으로 정보통신산업진흥원의 지원을 받아 수행된 연구임 (D0315-21-1002, XR기반 중증외상 처치훈련 시스템 구축)

REFERENCES

- [1] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces", in IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR), Nara, Japan, November 2007, pp. 225 - 23
- [2] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM: a versatile and accurate monocular SLAM system", IEEE Trans. Robot., 2015.
- [3] M. Bloesch, S. Omari, M. Hutter, and R. Siegwart, "Robust visual inertial odometry using a direct EKF-based approach", International Conference on Intelligent Robots and Systems(IROS), 2015.
- [4] Stefan Leutenegger, Simon Lynen, Michael Bosse, Roland Siegwart, and Paul Furgale. "Keyframe-based visual-inertial odometry using nonlinear optimization", The International Journal of Robotics Research, 2015.
- [5] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, "On-manifold preintegration for real-time visual-inertial odometry", IEEE Trans. Robot., vol. 33, no. 1, pp. 1 - 21, 2017.
- [6] D. Schubert, N. Demmel, L. von Stumberg, V. Usenko and D. Cremers, "Rolling-Shutter Modelling for Visual-Inertial Odometry", International Conference on Intelligent Robots and Systems(IROS), 2019
- [7] Jakob Engel, Vladlen Koltun, and Daniel Cremers. "Direct sparse odometry", IEEE transactions on pattern analysis and machine intelligence, 2017
- [8] Raul Mur-Artal and Juan D Tardos. "ORB-SLAM2: An open-source slam system for monocular, stereo, and rgb-d cameras", IEEE Transactions on Robotics, 2017.
- [9] R.Wang, M. Schworer, and D. Cremers. "Stereo DSO: Largescale direct sparse visual odometry with stereo cameras", In

- International Conference on Computer Vision (ICCV), 2017.
- [10] ARCore SDK site, <https://developers.google.com/ar>
- [11] ARKit SDK site, <https://developer.apple.com/kr/augmented-reality/arkit/>
- [12] Shing Yan Loo, Ali Jahani Amiri, Syamsiah Mashohor, Sai Hong Tang, and Hong Zhang. “CNN-SVO: Improving the mapping in semi-direct visual odometry using single image depth prediction”, International Conference on Robotics and Automation (ICRA), 2019
- [13] Clement Godard, Oisin Mac Aodha, Michael Firman, and Gabriel J. Brostow. “Digging into self-supervised monocular depth estimation”, The IEEE International Conference on Computer Vision (ICCV), 2019
- [14] Vuforia SDK site, <https://developer.vuforia.com/>
- [15] Vive Pro hardware spec site, <https://www.vive.com/kr/product/vive-pro-full-kit/>
- [16] D. Chekhlov, M. Pupilli, W. Mayol-Cuevas, and A. Calway. “Realtime and robust monocular SLAM using predictive multi-resolution descriptors”, In 2nd International Symposium on Visual Computing, 2006
- [17] E. Eade and T. Drummond. “Scalable monocular slam”, In Proc. IEEE Intl. Conference on Computer Vision and Pattern Recognition (CVPR), 2006
- [18] Jakob Engel, Jorg Stuckler, and Daniel Cremers. “Large-scale direct SLAM with stereo cameras”, In Intelligent Robots and Systems (IROS)



이 동 춘 (Lee, Dong Chun)

약 력 : 2001 경북대학교 컴퓨터공학 석사
2001-현재 한국전자통신연구원 연구원

관심분야 : 인공지능, 증강현실, 객체 인식, 가상현실



김 향 기 (Kim, Hang Kee)

약 력 : 2001 경북대학교 컴퓨터학과 석사
2001-현재 한국전자통신연구원 연구원

관심분야 : 인공지능, 가상현실, 객체 인식, 게임



이 기 석 (Lee, Ki Suk)

약 력 : 2001 성균관대학교 전기전자/컴퓨터공학 석사
2001-현재 한국전자통신연구원 연구원

관심분야 : 인공지능, 가상현실, 객체 인식, 증강현실,
물리 시뮬레이션

