JDCS
디지털콘텐츠학회논문지
Journal of Digital Contents Society
Vol. 24, No. 10, pp. 2509-2517, Oct. 2023

Check for updates

가상/증강현실 기반 원격 현장감(Telepresence)을 위한 시각 정보 전달 기술에 관한 연구

김 명 진[1,2*] · 김 현 수[3] · 김 현 종[3] · 황 찬 희[3]
[1*]전남대학교 ICT융합시스템공학과 조교수 [2*]전남대학교 바이오사이버네틱스연구센터 조교수
[3]전남대학교 컴퓨터정보통신공학과 학부생

Visual Information Transmission Methods for Virtual/Augmented-Reality – Based Telepresence

Myeongjin Kim[1,2*] · Hyeon-Su Kim[3] · Hyeon-Jong Kim[3] · Chan-Hue Hwang[3]

[1*]Assistant Professor, Department of ICT Convergence System Engineering, Chonnam National University, Gwangju 61186, Korea
[2*]Assistant Professor, Research Center for Biological Cybernetics, Chonnam National University, Gwangju 61186, Korea
[3]Undergraduate Student, Department of Computer Engineering, Chonnam National University, Gwangju 61186, Korea

[요 약]

원격 현장감(Telepresence)은 빠르게 발전하고 있는 기술로서, 사용자가 먼 환경에서 마치 직접 존재하는 것처럼 느낄 수 있게 해준다. 이러한 기술의 대표적인 응용 분야는 원격 회의, 원격 요양, 원격 조종 등이 있고 그 효용성이 검증되었다. 최근 들어 머리 착용형 디스플레이(HMD; head mounted display)가 발전함에 따라 가상/증강현실을 활용하여 원격 현장감을 향상하는 연구들이 수행되고 있다. 높은 원격 현장감을 위해서는 원격 환경에서 촬영된 영상 정보들을 분석하여 원격 환경과 원격 환경 내의 객체 정보를 실시간으로 사용자에게 전달해 주어야 한다. 본 논문에서는 가상/증강현실 기반 원격 현장감을 위한 원격 환경의 시각 정보를 HMD로 전송하기 위한 핵심 기술을 점 구름, 복셀(Voxel), View synthesis, 자세 추정 방법으로 분류하여 각 분야에서 주요 도전 과제와 최신 발전기술을 분석한다.

[Abstract]

Telepresence, a rapidly emerging technology, enables users to experience the sensation of being physically present in a distant environment. This sense of immersion has found utility in various domains ranging from remote conferencing and remote healthcare to distant control operations. One significant advancement that has augmented the experience of telepresence is the evolution of head-mounted displays (HMDs). There has been a burgeoning research trend that focuses on leveraging augmented reality (AR) and virtual reality (VR) to enhance remote operations and improve telepresence experiences. For achieving a heightened sense of telepresence, it is essential to analyse the video data captured from remote environments and stream both the surroundings and object details to the user in real-time. In this comprehensive study, we review the technologies that facilitate the transmission of VR/AR-based telepresence environments via HMDs. Research trends are categorized into point clouds, voxels, view synthesis, and pose estimation. The critical challenges faced by researchers in each category are explored and the latest research are introduced to provide an intuitive perspective to the readers.

색인어 : 가상현실, 증강현실, 원격 현장감, 원격 조작, 머리 착용형 디스플레이
Keyword : Virtual Reality, Augmented Reality, Telepresence, Teleoperation, Head Mounted Display

http://dx.doi.org/10.9728/dcs.2023.24.10.2509

Received 28 August 2023; **Revised** 20 September 2023
Accepted 04 October 2023

***Corresponding Author; Myeongjin Kim**
Tel: +82-62-530-1814
E-mail: myeongjin@chonnam.ac.kr

Ⅰ. 서 론

원격 현장감 (telepresence)이란 1980년대 초반 Marvin Minsky[1]에 의해 처음 소개된 개념으로 사용자가 먼 거리에 있는 특정 장소에 직접 가지 않고도 그 장소에 있는 것처럼 느끼게 해주는 일련의 기술을 의미한다. 원격 현장감 기술을 활용하면 시간과 공간의 제약에서 벗어나 원거리에 있는 다른 사람들과 같은 장소에 있는 것처럼 회의[2]-[4]를 할 수 있으며 원격 현장감을 높인 사업 회의(그림 1)의 효용성[4]이 기존의 오디오 혹은 오디오와 영상을 활용한 회의보다 높음을 보여주었다. 모니터와 센서를 탑재한 모바일 로봇(MRP; mobile robotic telepresence)[5](그림 1)을 활용하면 원거리에 있는 가족과 마주 보고 대화하듯이 대화를 할 수 있으며 실제 원격 환경에 있는 것처럼 돌아다닐 수 있어 도움이 필요한 가족을 원거리에서 요양을 할 수 있다. 몇몇 상용 제품들은 MRP의 키를 조정하여 앉거나 서는 동작을 모사할 수 있고 또한 주변 물체들을 잡고 움직이거나 감정을 표현하는 기능을 제공하여 더 높은 원격 현장감을 제공한다.

원격 현장감을 활용한 다른 응용 분야는 원격으로 로봇을 조종하는 원격 조종 (teleoperation)[6],[7](그림 1)이다. 원격 조종에서 원격 현장감 기술들을 적용하게 되면 사용자는 실제로 작업 환경에 있는 듯이 작업을 수행할 수 있어 작업의 효율이 올라가게 되어 산업 분야뿐만 아니라 다양한 분야에서 활용되고 있다.

그림 1. (a) 원격 현장감을 활용한 사업 미팅[4], (b) 가상/증강현실 기반 원격 현장감을 활용한 로봇 원격 조종[8], (c) 원격 현장감을 위한 모바일 로봇 [3]

Fig. 1. (a) Business meeting using telepresence[4], (b) VR/AR based robot teleoperation[8], (c) Mobile robot telepresence[3]

원격 현장감을 전달해주기 위해서는 물리적으로 떨어진 원격 환경에서 발생한 시각, 청각, 및 촉각 등을 사용자에게 전달해주어야 한다. 이를 위한 원격 환경의 시각, 청각 및 촉각 정보 측정 및 전달 기술이 필요하며 사용자를 위한 인터페이스가 필요하다. 시각 및 청각 피드백을 전달해주기 위한 인터페이스로써 기존에는 컴퓨터의 모니터와 스피커가 가장 많이 사용되었지만, 최근 들어 머리 착용형 디스플레이 (HMD; head mounted display) 기술[8]이 발달함에 사용자에게 더 높은 시야각과 해상도를 제공할 수 있게 되어 원격 현장감을 인터페이스로 활용되고 있다. 촉각을 전달해주는 대표적인 인터페이스로써, 펜 타입의 햅틱 장치가 사용되고 있으며, 다양한 사용자 착용형 햅틱 인터페이스 연구들이 활발히 진행되고 있다[9].

HMD를 통해 원격 환경을 시각화하여 전달하는 방법은 크게 원격 환경 정보를 토대로 원격 환경을 모사한 디지털 트윈을 사용자에게 가상현실(VR; virtual reality) 형태로 전달하는 방법과 사용자가 있는 현실 세계 위에 원격 환경의 정보로부터 생성된 가상 객체 혹은 영상을 덧씌워 증강현실 (AR; augmented reality) 형태로 전달하는 방법이 있다. 가상/증강현실을 통해 원격 환경의 시각 정보를 전달해주기 위해서는 원격 환경의 시각 정보를 측정 및 계산하기 위한 기술들이 필수적이다. 본 논문에서는 원격 환경 시각 정보를 측정 및 계산하여 사용자에게 전달해주는 기술들을 점 구름(point cloud), 복셀(voxel), view synthesis와 객체 자세 추정으로 분류하여 각 기술들의 특징과 문제점에 관해 설명하고 각 문제점을 해결하기 위해 수행된 연구를 소개한다.

Ⅱ. 원격 환경 시각 정보 전달 방법

2-1 점 구름 기반 방법

원격 환경을 전달하는 방식에는 크게 작업 중인 영역만을 전달하는 방식과 원격 환경 전체를 전달하는 방식[10]이 있다. 작업의 정확도는 두 경우 큰 차이를 보이지 않지만 작업 시간의 경우 원격 환경 전체를 보여줄 때 작업 시간이 줄어든다. 사용자가 원격 환경의 공간을 인식하기 위해서는 빛의 3원색으로 이루어진 2차원 RGB(reg, green, blue) 영상뿐만 아니라 깊이를 포함하는 3차원 공간 정보를 전달해주어야 한다. 이를 위한 대표적인 방법으로 공간의 깊이를 측정할 수 있는 RGBD(red, green, blue, and depth) 카메라 혹은 센서를 활용하여 원격 환경의 3차원 공간 정보를 3차원 점들의 집합인 점 구름 (point cloud)을 통해 전달해주는 것[11]이다.

점 구름은 3차원 (x, y, z) 정보를 가진 점들로 구성되어 있으며 점 들은 추가로 RGB 값을 가질 수 있다. 점 구름으로 전체 원격 환경을 전달해주는 방식은 높은 대역폭의 네트워크와 많은 계산을 해야 하는 문제가 있다. 이를 위해 점 구름 데이터에서 사전에 저장된 객체를 따로 처리하여 효율적으로

시각화하는 방법[12]이 제안되었으며 점 구름을 직접 전송하는 방식이 아닌 점 구름으로부터 3차원 모델 정보를 생성하여 3D 모델 정보를 전송하는 방법[13]이 제안되었다. 이를 통해 작업 속도 및 정확도를 향상했다. 3차원 모델 정보를 전송하게 되면 객체의 표면 정보를 메쉬(mesh)의 형태로 전송하면 되기 때문에 전송해야 할 데이터가 줄어드는 장점이 있다.

점 구름은 센서의 한계 때문에 노이즈가 발생하는 문제와 특정 영역의 표본이 부족하여 객체의 형상이나 특징을 제대로 나타내지 못하는 문제가 있다. 이러한 문제를 해결하기 위해 점 구름 복원(point cloud restoration) 방법들이 제안되었다. 점 구름 복원 방법은 크게 노이즈를 제거하는 denosing 방법[14],[15]과 표본이 부족한 영역에서 추가로 점들을 생성하는 upsampling 방법[15],[16]으로 나뉘며 문제 해결 방식에 따라 최적화 기반 방법과 인공지능 기반 방법으로 분류할 수 있으며 최근 들어서는 인공지능 기반 방법들이 중점적으로 연구되고 있다. 점 구름 데이터에 있는 노이즈를 제거하기 위해 노이즈에 의한 발생한 점의 변위를 네트워크 모델로 예측한 뒤에 노이즈에 의한 변위를 제거하는 방법[17],[18]이 제안되었다. 이러한 방법의 문제는 과대 혹은 과소 추정된 변위로 인해 점 구름으로 표현된 물체의 형상이 왜곡되는 문제가 있다. 점의 위치 보정을 위해 변위를 추정하여 적용하는 대신 보정 방향을 추정한 뒤 점진적으로 조금씩 점들을 이동시켜 노이즈를 제거하는 방법[19](그림 2)이 제안되었다.

점 구름 데이터만을 활용해 원격 환경을 그대로 전달해주게 되면 사용자가 환경에서 객체를 구분하기 어려운 단점이 있다. 이러한 문제를 해결하기 위해 점 구름 데이터에서 객체를 찾아서 분류하기 위한 연구들이 진행되었다. 2017년에 처음 제안된 PointNet 방법[20]은 불규칙한 점 구름 데이터부터 해당 객체를 분류(classification)하거나, 하나의 객체를 여러 부분으로 나누는 부분 분할(part segmentation)을 할 수 있으며, 여러 객체가 있는 환경에서 각 객체를 분류할 수 있다. 불규칙한 점 구름 데이터를 처리하기 위해 입력 순서에 상관없이 데이터를 처리하는 정규화 네트워크 구조를 제안하였다. PointNet은 각 점을 독립적으로 처리하여 전체 데이터에서 특징을 추출하여 국소적인 특징을 분석하는 데 한계가 있다. 이러한 문제를 해결하기 위해 PointNet++[21](그림 3)가 도입되었다. PointNet++는 계층적인 접근 방식을 사용하여 점 구름 데이터를 처리한다. 우선 작은 군집에서 PointNet을 적용하여 특징을 추출하고 이들의 결과를 통합한다. PointNet++은 이 과정을 여러 계층에서 수행하여 PointNet보다 국소적인 특징들을 더 잘 분석할 수 있다. 이와 유사한 방법으로 DDGCN(graph convolutional neural network architecture) 방법[22]이 제안되었다. DDGCN 방법에서는 k-NN(k-nearest neighbor)방법을 통해 인접한 점들끼리 묶은 뒤 그룹별로 특징을 추출하고 추출한 특징들을 graph convolutional neural network를 통해 통합하였다. Graph와 같은 계층구조를 활용하므로 더 복잡한 환경을 정밀하게 분석할 수 있는 장점이 있다.

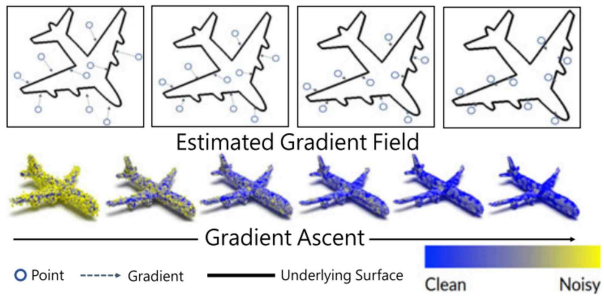

그림 2. 인공지능 기반의 점 구름 복원 방법[19]
Fig. 2. Neural network based point cloud restoration[19]

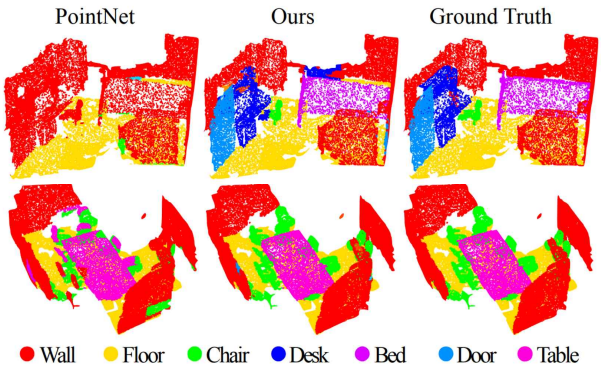

그림 3. PointNet, PointNet++(Ours)를 활용한 점 구름 데이터의 객체 분류[21]
Fig. 3. Classification of point cloud data using PointNet and PointNet++(Ours)[21]

2-2 복셀 기반 방법

점 구름 데이터는 원격 환경의 3차원 공간 정보를 전달해 줄 수 있어 유용하지만 불규칙한 점 데이터를 활용하므로 데이터 저장 및 처리 효율이 떨어지는 단점이 있다. 이를 위해 점 구름으로 측정된 원격 환경의 3차원 공간 정보를 균일한 정육면체 형상의 3차원 화소, 복셀로 변환하여 복셀맵을 만드는 방법[23]-[25]이 활용되고 있다. 복셀 정보를 그대로 시각화하여 전달해주거나 혹은 복셀로부터 메쉬를 재구성하여 사용자에게 시각화해 전달해줄 수 있다.

점 구름 대신 복셀맵을 활용하면 데이터 구조를 단순화하여 메모리 효율을 높일 수 있고, 점 구름 데이터의 노이즈를 줄일 수 있다. 또한 복셀맵 정보를 시각화할 때도 점 구름 데이터에 비해 높은 품질의 이미지를 얻을 수 있어 원격 현장감의 현실감을 높일 수 있다.

복셀맵의 메모리 효율성과 원격 환경 변화에 따른 업데이트 효율성을 증가시키기 위한 연구들이 수행되고 있다. Octomap 방법[26](그림 4)은 복셀맵의 효율을 증가시키기 위해 기존의 균일한 복셀로 전체 공간을 모델링 하는 대신 계층구조의 octree라는 계층구조를 활용하여 큰 3차원 복셀을 8등분 하여 계층적으로 나누어 사용한다. 이렇게 하게 되면 3차원 공간에

있는 물체들을 그 크기에 적합한 복셀로 모델링 할 수 있어 더 적은 복셀을 사용하여 3차원 공간을 표현할 수 있어 같은 메모리를 가지고 더 넓은 공간과 많은 객체를 표현할 수 있다. 또한 각 복셀내 객체의 존재 여부를 확률론적으로 계산함으로써 원격 환경의 동적인 변화에도 효율적으로 대처할 수 있다. 제안된 방법은 균일 복셀맵 대비 44% 향상된 메모리 효율을 보여주었다. Octomap 방법에서 동적인 물체의 인식률을 높이기 위해 시간에 따른 변화를 예측하는 RNN(recurrent neural network) 모델이 적용된 Reccurent-octomap[27]이 제안되었다. 제안된 방법은 기존 octomap 대비 자동차, 보행자, 자전거 등의 동적 물체를 더 잘 표현한다. 기존의 octomap은 복셀 내 객체의 존재 여부만을 판단하여 공간을 모델링 하여 측정되지 않은 영역에 대해서는 명시적인 모델링이 불가능한 단점이 있다. 이를 해결하기 위해 UFOmap[28]에서는 객체 존재 여부 외 불분명정보까지 포함한 3가지 지표를 활용하여 공간을 모델링 하였고, 이전 방법과 달리 모든 영역을 3가지 지표로 균일하게 처리 함으로써 메모리 효율을 3배 향상했다. 점 구름 데이터를 복셀맵으로 변환시킬 때 점 구름 데이터의 노이즈가 포함되어 부정확하게 변환되는 문제가 발생한다. 이를 해결하기 위해 k-NN(k-nearest neighbor)를 활용해 인접한 점이 많은 점들만을 변환하여 노이즈를 제거하는 방법[29]이 개발되었다. 이 외에도 메모리 사용량을 줄이면서도 더 넓은 영역을 실시간으로 모델링하는 연구들이 수행되고 있다.

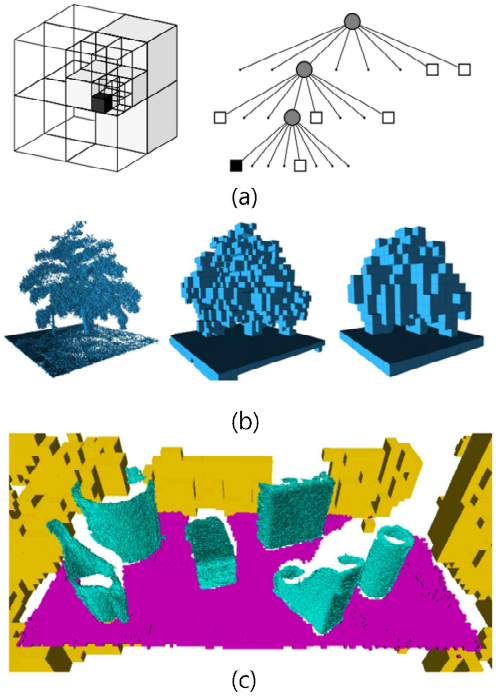

그림 4. (a) 복셀의 계층구조, (b) 복셀 계층구조의 Level에 따른 복셀의 크기, (c) 여러 계층 구조로 나타낸 복셀맵[26]

Fig. 4. (a) the hierarchical structure of voxels, (b) the size of voxels according to the level, (c) Voxel map represented by multiple hierarchical structures[26]

2-3 View Synthesis 기반 방법

View synthesis는 다양한 각도와 위치에서 캡처된 2차원 이미지를 바탕으로, 임의의 위치와 각도에서 촬영된 2차원 이미지를 추정하는 기술을 가리킨다. 이 기술을 활용함으로써, 원격 환경에서 촬영된 2차원 이미지 및 비디오를 이용하여, 원격 환경 임의의 위치와 각도에 대한 2차원 이미지를 생성할 수 있다. 이는 사용자가 원격 환경 내에서 자유롭게 이동하며 다양한 위치와 각도에서 원격 환경을 관찰하는 것을 가능하게 한다. 인공지능 기술의 발전에 따라, View synthesis 분야에서도 인공지능을 기반으로 한 기술이 개발되고 있다. NeRF(neural radiance fields)[30]는 2020년에 제안된 인공지능 기반의 view synthesis 모델로서, 다수의 2D 이미지를 입력으로 받아, 임의의 관점에서의 이미지를 생성하는 기술을 제시하였다.

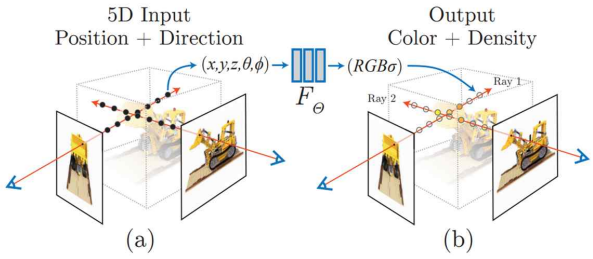

그림 5. Neural Radiance Fields 방법의 아이디어[29] (a) 주어진 카메라의 위치 및 각도에 따른 ray 생성, (b) 인공지능 기반 ray가 지나는 복셀의 RGB 및 밀도 값 추정

Fig. 5. The concept idea of Neural Radiance Fields[29] (a) ray by the given camera position and orientation, (b) neural network based estimation of voxel RGB and density

NeRF의 아이디어는 고전적인 볼륨 렌더링 기법을 활용하는 것이다. 고전적인 볼륨 렌더링은 RGB 및 밀도 정보를 포함하는 3차원 배열 데이터인 복셀을 이용하여 특정 공간의 이미지를 생성하는 기술이다. 이 기술은 카메라로부터 복셀을 향하는 ray를 생성하고, 이 ray가 통과한 복셀의 값의 표본을 뽑아 카메라 시점에서 본 2차원 이미지를 생성한다. NeRF는 다수의 2차원 이미지를 통해 역으로 이미지 생성을 위한 3차원 공간의 복셀 값을 추정한다. 고전적인 볼륨 렌더링에서는 이미지 생성을 위한 정보가 실제 복셀에 저장되지만, NeRF에서는 이 정보가 실제 복셀 대신 네트워크 모델의 MLP(multi layer perceptron)에 저장된다. 이를 통해 NeRF는 추정된 복셀 값들을 활용해 임의의 시점에서 바라본 2차원 이미지 생성이 가능하다.

그러나 NeRF는 학습에 하루 이상이 소요되며, 2차원 이미지를 렌더링하는데도 상당한 시간이 소요되고 렌더링할 수 있는 영역이 한정되고, 특정 영역에서 이미지 품질이 떨어지

는 한계가 있다. 이러한 문제점을 극복하기 위해 다양한 연구가 수행되었다. 2020년에는 카메라로부터 멀리 떨어진 배경 영역의 이미지 렌더링 품질을 개선하기 위해 NeRF++[31]가 제안되었다. 이 방법은 가까운 영역과 먼 영역을 정의하고, 각 영역에 대해 별도의 매개변수를 설정함으로써 배경 영역의 이미지 렌더링 품질을 향상했다. NeRF는 이미지의 각 픽셀에 대한 RGB 값을 개별적으로 계산하므로 계단 현상(aliasing)이 발생한다. 이를 해결하기 위해 기존의 ray 대신에 원뿔형의 frustum을 사용하여 거리에 따라 표본 영역을 증가시키는 mip-NeRF[32] 방식이 제안되었다. 이 방식은 이미지 렌더링 품질을 향상하는 동시에, 연산 속도를 기존대비 7% 향상했다. NeRF 계열의 방법들은 한정된 영역의 렌더링에 중점을 두는 방식이므로, 넓은 영역을 렌더링 하는 데는 근본적인 한계가 있다. 이러한 문제점을 극복하기 위해 mip-NeRF 360[33]이 제안되었다. mip-NeRF 360은 렌더링 되는 영역 외부의 주변 영역을 정규화하여 렌더링 영역으로 압축하는 방식을 사용하였다. 이를 통해 렌더링 한정된 영역이 아닌 넓은 영역을 렌더링할 수 있게 되었으며, 렌더링 영역의 압축을 통해 학습 속도를 300% 향상했다.

최근에는 해시 인코딩과 선형 보간을 활용하여 데이터 접근 속도를 빠르게 하고, 네트워크 구조를 단순화하여 NeRF와 같은 view synthesis의 속도를 향상하는 iNGP(instant neural graphics primitives)[34]이 제안되었다. 그리고 mip-NeRF 360과 iNGP의 강점을 결합하여 넓은 영역을 빠른 속도로 렌더링할 수 있는 zip-NeRF[35]가 제안되었다. 또한 zip-NeRF에서는 추가적인 비용함수를 도입하여 이미지 생성 시 z축을 따라 특정 영역이 렌더링 되지 않고 사라지는 z-aliasing 문제를 개선하였다.

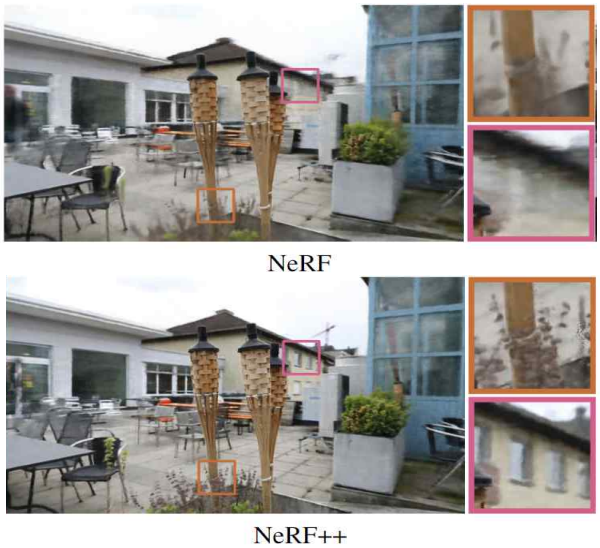

그림 6. NeRF와 NeRF++로 생성된 2차원 이미지의 해상도 비교[31]

Fig. 6. Comparison of image resolution generated by NeRF and NeRF++[31]

그 외에도 어두운 환경에 특화된 NeRF in the Dark[36], 빛 반사에 특화된 Ref-NeRF[37], 움직이는 사람에 특화된 HumanNeRF[38] 등과 같이 특정 분야에 특화된 NeRF도 연구되고 있으며 속도 개선, 입력 이미지 수 축소, 카메라 자세 추정, 넓은 범위의 장면 렌더링 등을 목표로 하는 다양한 NeRF 개선 연구들이 지속해서 이루어지고 있으며, 이는 현재도 계속 진행 중이다. 앞으로 높은 원격 현장감을 위해 실시간으로 원격 환경을 전달해 주는 다양한 NeRF 기반 방법들이 개발될 것으로 기대된다.

2-4 객체 자세 추정 방법

원격 환경내 객체들의 정보를 전달하는 또 다른 방법은 원격 환경내 객체와 동일한 3차원 모델을 준비한 뒤 원격 환경에서의 객체의 3병진, 3회전 자유도를 추정하여 가상/증강현실내의 동일한 해당 객체의 자세가 같아지도록 하는 것이다. 이러한 자세 추정 기술은 원격 현장감에 활용될 수 있을 뿐만 아니라 로봇이 객체를 인식할 수 있게 하므로 로봇 자동 조작에도 활용될 수 있다.

원격 환경 내 여러 객체가 있을 경우 객체 검출 및 분류 알고리즘을 함께 사용하여 객체를 구분하고 객체 별로 자세를 추정하는 방식으로 활용되고 있다. 2차원 단일 이미지를 활용하여 객체를 추정하기 위해 PoseCNN(convolutional neural network for 6D object pose estimation)[39]이 제안되었다. 객체 분류와 객체의 위치, 회전 추정을 다른 문제로 구분하여 해결한 방법으로 이미지를 이루는 픽셀을 각 객체별로 분류한 뒤, 각 객체의 중심을 추정하여 병진 3자유도 위치를 추정한다. 그리고 물체를 감싸는 2차원 사각형 형상의 bounding box를 생성하여 물체의 회전 3자유도를 추정한다. PoseCNN 방법은 2차원 단일 이미지로부터 각 객체의 6자유도를 추정하는 방법이지만 오차가 크다는 단점이 있다. 이를 보완하기 위해 DOPE(deep object pose estimation)[40]라는 방법이 제안되었다. DOPE는 VGG-19[41]를 활용하여 객체를 분류하고 객체를 감싸는 bounding box를 추정하여 6자유도 위치를 추정한다. 기존 PoseCNN과 달리 2차원 사각형 대신 3차원 직육면체의 bounding box를 사용하여 6자유도 자세 추정 정확도를 향상했다. 또한, 다양한 배경 이미지와 합성된 훈련 데이터를 생성하여 훈련에 활용함으로써 부족한 데이터 문제를 해결하였고 방법의 강인성을 향상했다.

위 두 방법과는 다르게 이미지가 아닌 다른 데이터를 활용하여 객체의 자세를 추정하는 방법들도 제안되었다. PPR-Net(point-wise pose regression network)[42]은 점 구름 데이터를 활용하여 객체를 분류하고 자세를 추정하였다. 가시성 변수를 추가하여 객체의 구분이 모호하거나 자세를 추정하기에 어려운 경우 해당 객체는 배제하여 여러 객체가 겹쳐 있는 경우에 대해 추정오차를 줄였다. OP-NET(object pose estimation network)[43]은 깊이(depth) 데이터만을 가지고 물체의 자세를 추정하는 방식을 제안하였다. 사전에 가지고 있

던 객체의 3차원 모델을 활용하여 학습데이터를 생성하였다. 하지만 깊이를 활용하는 방식은 여러 객체가 유사한 깊이 값을 가진 경우에 오차가 큰 단점이 있다. 단일 이미지를 활용하여 객체의 자세를 결정하는 방식들은 객체를 검출하고 자세를 추정하는 과정이 분리되어 있어 이를 하나의 네트워크에서 처리하고자 GDR-NET(geometry-guided direct regression network)[44]이 제안되었다. 자세를 추정하고자 하는 객체의 3차원 모델을 활용하여 객체의 표면에 대한 정보와 객체의 2D-3D 사상에 대한 데이터를 추출하여 자세를 추정한다.

기존의 자세 추정 방법들은 지도학습 기반으로 학습을 위해서는 각 객체의 6자유도에 대한 ground truth 값을 제공해 주어야 하는 한계가 있다. 이를 해결하고자 UDA-COPE (unsupervised domain adaptation for category-level object pose estimation)[45]가 제안되었다. RGBD 데이터를 2D와 3D로 분할 추출하고 이를 선생-학생 네트워크 구조를 활용하여 비지도 기반으로 학습하였고, 그 결과 지도 기반과 유사한 정확도를 달성하였다.

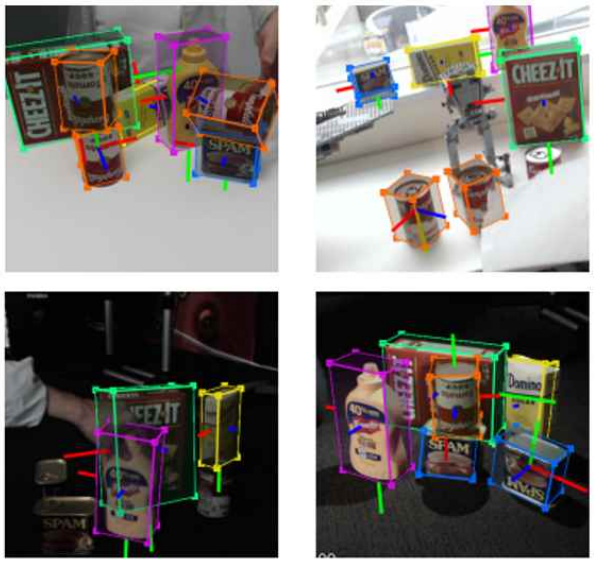

그림 7. 직육면체 bounding box를 활용한 DOPE[40]의 객체의 6자유도 위치 추정

Fig. 7. 6-DOF pose estimation of the object using a cuboid bounding box in DOPE[40]

Ⅲ. 기존 방법 비교 분석

본 논문에서 조사한 원격 환경 시각 정보 전달 방법들을 원격 환경으로부터 측정하는 정보, 전처리 여부, 시각화된 정보의 질 및 실시간 업데이트 성능 및 기타 제공할 수 있는 정보 측면에서 비교하여 표1로 정리하였다. 점 구름 데이터를 전달하는 방식은 RGBD 카메라로 촬영된 정보를 바로 전달할 수 있어 전처리가 필요 없고 실시간으로 변하는 원격 환경을 바로 전달해 줄 수 있지만 센서의 한계로 노이즈가 발생하여 이를 제거하기 위한 연구들이 수행되고 있다. 점 구름 방식은 점 구름 형태로 객체를 시각화하여 사용자에게 보여주므로 시각 정보의 질이 다소 떨어지는 단점이 있다. 점 구름 데이터로부터 공간 내 객체를 인식 및 분류하는 연구들도 인공지능 기술이 발전함에 따라 많은 연구가 이루어졌으며 점 구름을 통한 객체 인식 방법은 최근 자세 추정 방법에도 활용되고 있다.

복셀맵을 활용해 원격 환경의 공간 정보를 전달하는 방법은 점 구름 데이터가 가진 노이즈를 제거하고 계층구조를 통해 효율적으로 데이터를 관리 할 수 있지만 점 구름으로부터 데이터를 변환하는 단계가 필요한 단점이 있다. 시각 정보의 경우 복셀로부터 3차원 표면 모델을 추출하는 방법을 활용하면 시각 정보의 질을 높일 수 있을 것이다. 실시간으로 변하는 원격 환경을 위해 계층구조를 효율적으로 업데이트하는 연구가 이루어져야 할 것이다.

View synthesis 기반 방법은 실제 사용자가 원격 환경에 있는 것처럼 RGB 이미지를 볼 수 있지만 학습에 시간이 오래 걸리고 정적인 원격 환경에만 적용할 수 있다는 단점이 있다. 최근에는 학습 시간을 크게 단축한 방법들이 제시되고 있다. 하지만 아직 실시간으로 변하는 동적인 원격 환경을 렌더링하는 것은 숙제로 남아 있다.

촬영된 영상을 통해 환경 내 객체를 실시간으로 추정하는 방법은 추정된 자세로부터 3차원 모델의 자세를 결정한 뒤 3차원 모델로부터 시각 정보를 생성하여 전달하므로 높은 현실감을 얻을 수 있다. 또한 객체의 자세 정보는 로봇 제어에 활용되면 큰 이점을 가질 수 있으므로 로봇 분야에서 활발히 연구되고 있다. 자세 추정 방법은 학습데이터가 존재하는 객체만을 인식한다는 한계가 있지만 이러한 한계를 극복하기 위해 가상공간에서 원하는 객체 형상에 대해 학습데이터를 생성하는 방법과 비지도 기반 학습 방법들도 연구되고 있다.

표 1. 원격 환경 시각 정보 전달 방법의 활용 분야, 필요정보, 전처리, 전달되는 시각 정보, 기타 정보, 실시간 업데이트 성능 비교 분석

Table 1. Comparative analysis of availability, necessary information, preprocessing, transmitted visual information, etc information, and real-time performance in methods transmitting visual information of remote environment

방법	활용 분야	필요 정보	전처리	시각 정보	기타 정보	실시간 성능
점 구름	AR,VR	RGBD	불필요	점으로 구성된 3차원 모델	객체 분류	빠름
복셀	AR,VR	RGBD	필요	복셀로 구성된 3차원 모델	객체 분류	중간
View synthesis	VR	RGB, RGBD	필요	2차원 RGB 이미지	-	느림
자세 추정	AR,VR	RGB, RGBD	필요	3차원 표면 객체 모델	객체 분류, 객체 자세	빠름

Ⅳ. 결 론

최근 들어 HMD 장비의 계산 성능과 사용성이 개선됨에 따라 HMD를 활용하여 원격 환경을 가상/증강현실을 통해 사용자에게 전달하여 원격 현장감을 제공해주는 연구들이 수행되고 있다. 본 논문에서는 HMD를 통한 원격 환경을 전달하기 위한 핵심 기술인 점 구름, 복셀, view synthesis, 자세 추정 기술들을 최근 기술을 분석하였다.

점 구름 방법은 시각 정보의 질이 낮은 단점이 존재하고 그 외 공간에 대한 정보들을 가공하여 사용자에게 전달하는 방법들은 여전히 처리 속도, 정확도, 실용성 등에서의 한계가 존재한다. 하지만 가상/증강현실을 통한 원격 환경 전달의 연구와 기술은 인공지능 기술이 발전함에 따라 지속해서 발전하고 있으며, 원격 환경을 실시간으로 전달하여 현실감 높은 원격 현장감 경험을 제공할 가능성이 열리고 있다. 또한 사용자에게 원격 현장의 햅틱 감각을 전달해 줄 수 있는 착용형 햅틱 슈트와 같은 햅틱 인터페이스 기술이 뒷받침된다면 실제와 같은 원격 현장감을 제공해 줄 수 있을 것이다.

감사의 글

본 연구는 과학기술정보통신부의 한국연구재단(No. 2022R1F1A1072309), 과학기술정보통신부 및 정보통신기획평가원의 지역지능화혁신인재양성사업(IITP-2023-RS-2022-00156287) 및 산업통상자원부 및 한국산업기술기획평가원(No. 1415187415, No. 20025750)의 지원으로 수행 되었음.

참고문헌

[1] M. Minsky, *Telepresence*, 1980.

[2] W. Standaert, S. Muylle, and A. Basu, “An Empirical Study of the Effectiveness of Telepresence as a Business Meeting Mode,” *Information Technology and Management*, Vol. 17, No. 4, pp. 323-339, December 2016. https://doi.org/10.1007/S10799-015-0221-9

[3] A. Orlandini et al. “Excite Project: A Review of Forty-Two Months of Robotic Telepresence Technology Evolution,” *Presence: Teleoperators and Virtual Environments*, Vol. 25, No. 3, pp. 204-221, 2016. https://doi.org/10.1162/Pres_A_00262

[4] V. A. Nguyen Et Al. “Item: Immersive Telepresence For Entertainment and Meetings—A Practical Approach,” In *IEEE Journal of Selected Topics In Signal Processing*, Vol. 9, No. 3, pp. 546-561, April 2015. https://doi.org/10.1109/Jstsp.2014.2375819

[5] Kristoffersson, A., Coradeschi, S., and Loutfi, A., “A Review of Mobile Robotic Telepresence,” *Advances in Human-Computer Interaction*, Vol. 2013, p. 17, 2013. https://doi.org/10.1155/2013/902316

[6] M. Wonsick and T. Padir, “A Systematic Review of Virtual Reality Interfaces for Controlling and Interacting with Robots,” *Applied Sciences*, Vol. 10, No. 24, pp. 9051, December 2020. https://doi.org/10.3390/App10249051

[7] F. Brizzi et al., “Effects of Augmented Reality on the Performance of Teleoperated Industrial Assembly Tasks in a Robotic Embodiment,” In *IEEE Transactions on Human-Machine Systems*, Vol. 48, No. 2, pp. 197-206, April 2018. https://doi.org/10.1109/Thms.2017.2782490

[8] C. Chang, et al., “Toward the Next-Generation VR/AR Optics: A Review of Holographicnear-Eye Displays from a Human-Centric Perspective,” *Optica*, Vol. 7, pp. 1563-1578, 2020. https://doi.org/10.1364/Optica.406004

[9] G. S. Giri, Y. Maddahi, and K. Zareinia, “An Application-Based Review of Haptics Technology,” *Robotics*, Vol. 10, No. 1, pp. 29, February 2021. https://doi.org/10.3390/Robotics10010029

[10] D. B. Van De Merwe et al. “Human-Robot Interaction During Virtual Reality Mediated Teleoperation: How Environment Information Affects Spatial Task Performance and Operator Situation Awareness,” *Virtual, Augmented and Mixed Reality. Applications and Case Studies: 11th International Conference, HCII 2019*, Orlando, FL, pp. 163-177, 2019. https://doi.org/10.1007/978-3-030-21565-1_11

[11] Y. H. Su et al., “Development of an Effective 3D VR-Based Manipulation System for Industrial Robot Manipulators,” *2019 12th Asian Control Conference (ASCC)*, Kitakyushu, Japan, pp. 1-6, 2019.

[12] S. Kohn et al. “Towards a Real-Time Environment Reconstruction for VR-Based Teleoperation Through Model Segmentation,” *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Madrid, Spain, pp. 1-9, 2018. https://doi.org/10.1109/Iros.2018.8594053

[13] F. Brizzi, L. Peppoloni, A. Graziano, E. Di Stefano, C. A. Avizzano, and E. Ruffaldi, “Effects of Augmented Reality on the Performance of Teleoperated Industrial Assembly Tasks in a Robotic Embodiment,” In *IEEE Transactions on Human-Machine Systems*, Vol. 48, No. 2, pp. 197-206, April 2018. https://doi.org/10.1109/Thms.2017.2782490

[14] E. Mattei and A. Castrodad, “Point Cloud Denoising via Moving RPCA,” *Computer Graphics Forum*, Vol. 36, No. 8, pp. 123-137, 2017. https://doi.org/10.1111/Cgf.13068

[15] H. Huang, S. Wu, M. Gong, D. Cohen-Or, U. Ascher, and H. Zhang, “Edge-Aware Point Set Resampling,” *ACM*

Transactions on Graphics (ToG), Vol. 32, No. 1, pp. 1-12, 2013. https://doi.org/10.1145/2421636.2421645

[16] Y. Lipman et al., "Parameterization-Free Projection for Geometry Reconstruction," *ACM Transactions on Graphics (ToG)*, Vol. 26, No. 3, pp. 22-es, 2007. https://doi.org/10.1145/1276377.1276405

[17] M. J. Rakotosaona et al., "Pointcleannet: Learning to Denoise and Remove Outliers from Dense Point Clouds," *Computer Graphics Forum,* Vol. 39, No. 1, pp. 185-203, 2020. https://doi.org/10.1111/Cgf.13753

[18] F. Pistilli et al., "Learning Graph-Convolutional Representations for Point Cloud Denoising," *European Conference on Computer Vision*, Cham: Springer International Publishing, 2020. https://doi.org/10.1007/978-3-030-58565-5_7

[19] H. Chen, B. Du, S. Luo, and W. Hu, "Deep Point Set Resampling via Gradient Fields," In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 45, No. 3, pp. 2913-2930, March 2023. https://doi.org/10.1109/Tpami.2022.3175183

[20] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, "Pointnet: Deep Learning on Point Sets for 3D Classification and Segmentation," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, pp. 77-85, 2017. https://doi.org/10.1109/Cvpr.2017.16

[21] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space," *Advances in Neural Information Processing Systems 30(NIPS 2017)*, 2017.

[22] L. F. Chen and Q. Zhang, "DDGCN: Graph Convolution Network Based on Direction and Distance for Point Cloud Learning," *Visual Computer*, Vol. 39. No. 3 pp. 863-873, 2023. https://doi.org/10.1007/S00371-021-02351-8

[23] J. Papon, A. Abramov, M. Schoeler, and F. Wörgötter, "Voxel Cloud Connectivity Segmentation - Supervoxels for Point Clouds," *2013 IEEE Conference on Computer Vision and Pattern Recognition*, Portland, OR, pp. 2027-2034, 2023. https://doi.org/10.1109/Cvpr.2013.264

[24] I. Dryanovski, W. Morris, and J. Xiao, "Multi-Volume Occupancy Grids: An Efficient Probabilistic 3D Mapping Model for Micro Aerial Vehicles," *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Taipei, Taiwan, pp. 1553-1559, 2010. https://doi.org/10.1109/Iros.2010.5652494

[25] B. Douillard et al., "Hybrid Elevation Maps: 3D Surface Models for Segmentation," *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Taipei, Taiwan, pp. 1532-1538, 2010. https://doi.org/10.1109/Iros.2010.5650541

[26] A. Hornung et al., "Octomap: An Efficient Probabilistic 3D Mapping Framework Based on Octrees," *Autonomous Robots*, Vol. 34, No. 3, pp. 189-206, 2013. https://doi.org/10.1007/S10514-012-9321-0

[27] L. Sun et al., "Recurrent-Octomap: Learning State-Based Map Refinement for Long-Term Semantic Mapping with 3-D-Lidar Data," in *IEEE Robotics and Automation Letters*, Vol. 3, No. 4, pp. 3749-3756, October 2018. https://doi.org/10.1109/Lra.2018.2856268

[28] D. Duberg and P. Jensfelt, "UFOMap: An Efficient Probabilistic 3D Mapping Framework That Embraces the Unknown," in *IEEE Robotics and Automation Letters*, Vol. 5, No. 4, pp. 6411-6418, October 2020. https://doi.org/10.1109/Lra.2020.3013861

[29] Y. Miao, A. Hunter, and I. Georgilas, "An Occupancy Mapping Method Based on K-Nearest Neighbours," *Sensors*, Vol. 22, No. 1, 139, 2021. https://doi.org/10.3390/S22010139

[30] B. Mildenhall et al., "NERF: Representing Scenes as Neural Radiance Fields for View Synthesis," *Communications of the ACM*, Vol. 65, No. 1, pp. 99-106, 2021. https://doi.org/10.1145/3503250

[31] K. Zhang et al., "Nerf++: Analyzing and Improving Neural Radiance Fields," *Arxiv Preprint Arxiv:2010.07492*, 2020. https://doi.org/10.48550/Arxiv.2010.07492

[32] J. T. Barronet et al., "MIP-NERF: A Multiscale Representation for Anti-Aliasing Neural Radiance Fields," *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada, pp. 5835-5844, 2021. https://doi.org/10.1109/Iccv48922.2021.00580

[33] J. T. Barronet et al., "MIP-NERF 360: Unbounded Anti-Aliased Neural Radiance Fields," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, pp. 5460-5469, 2022. https://doi.org/10.1109/Cvpr52688.2022.00539

[34] T. Müller et al., "Instant Neural Graphics Primitives with a Multiresolution Hash Encoding." *ACM Transactions on Graphics (ToG)*, Vol. 41, No. 4, 102, 2022. https://doi.org/10.1145/3528223.3530127

[35] J. T. Barron et al., "Zip-Nerf: Anti-Aliased Grid-Based Neural Radiance Fields," *Arxiv Preprint Arxiv:2304.06706*, 2023. https://doi.org/10.48550/Arxiv.2304.06706

[36] B. Mildenhall et al., "Nerf in the Dark: High Dynamic Range View Synthesis From Noisy Raw Images," *2022 IEEE/CVF Conference on Computer Vision and Pattern*

Recognition (CVPR), New Orleans, LA, pp. 16169-16178, 2022. https://doi.org/10.1109/Cvpr52688.2022.01571

[37] D. Verbin et al., "Ref-Nerf: Structured View-Dependent Appearance for Neural Radiance Fields," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, pp. 5481-5490, 2022. https://doi.org/10.1109/Cvpr52688.2022.00541

[38] C. Weng et al., "Humannerf: Free-Viewpoint Rendering of Moving People from Monocular Video," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, pp. 16189-16199, 2022. https://doi.org/10.1109/Cvpr52688.2022.01573.

[39] Y. Xiang et al., "Posecnn: A Convolutional Neural Network for 6D Object Pose Estimation in Cluttered Scenes," *Arxiv Preprint Arxiv:1711.00199*, 2017. https://doi.org/10.48550/Arxiv.1711.00199

[40] J. Tremblay et al., "Deep Object Pose Estimation for Semantic Robotic Grasping of Household Objects," *Arxiv Preprint Arxiv:1809.10790*, 2018. https://doi.org/10.48550/Arxiv.1809.10790

[41] K. Simonyan and Z. Andrew, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *Arxiv Preprint Arxiv:1409.1556*, 2014. https://doi.org/10.48550/Arxiv.1409.1556

[42] Z. Dong et al., "PPR-Net:Point-Wise Pose Regression Network for Instance Segmentation and 6D Pose Estimation In Bin-Picking Scenarios," *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Macau, China, pp. 1773-1780, 2019. https://doi.org/10.1109/Iros40897.2019.8967895.

[43] K. Kleeberger and M. F. Huber, "Single Shot 6D Object Pose Estimation," *2020 IEEE International Conference on Robotics and Automation (ICRA)*, Paris, France, pp. 6239-6245, 2020. https://doi.org/10.1109/Icra40945.2020.9197207.

[44] G. Wang, F. Manhardt, F. Tombari, and X. Ji, "GDR-Net: Geometry-Guided Direct Regression Network for Monocular 6D Object Pose Estimation," *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, pp. 16606-16616, 2021. https://doi.org/10.1109/Cvpr46437.2021.01634.

[45] T. Lee et al., "Uda-Cope: Unsupervised Domain Adaptation for Category-Level Object Pose Estimation," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, pp. 14871-14880, 2022. https://doi.org/10.1109/CvPR52688.2022.01447.

김명진(Myeongjin Kim)

2010년 : 부산대학교 기계공학부(학사)
2012년 : 한국과학기술원 기계항공시스템공학부(석사)
2019년 : 한국과학기술원 기계공학과 (박사)
2019년: 한국과학기술원 연수연구원
2019년～2021년: Imperial College London, Research Associate
2022년～현 재: 전남대학교 컴퓨터정보통신공학과 조교수
※관심분야 : 가상현실(Virtual reality), 증강현실(Augmented reality), 햅틱(Haptic)

김현수(Hyeon-Su Kim)

2019년～현재: 전남대학교 컴퓨터정보통신공학과(학사)
2022년～현 재: VHMR 학부연구생
※관심분야 : 컴퓨터 비전(Computer vision), 혼합현실(Mixed reality), 시뮬레이션(Simulation), 인공지능

김현종(Hyeon-Jong Kim)

2019년～현재: 전남대학교 컴퓨터정보통신공학과(학사)
2022년～현 재: VHMR 학부연구생
※관심분야 : 컴퓨터 비전(Computer vision), 혼합현실(Mixed reality), 시뮬레이션(Simulation), 인공지능

황찬희(Chan-Hue Hwang)

2020년～현재: 전남대학교 컴퓨터정보통신공학과(학사)
2023년～현 재: VHMR 학부연구생
※관심분야 : 컴퓨터 비전(Computer vision), 혼합현실(Mixed reality), 시뮬레이션(Simulation), 인공지능