

웨어러블 AR 기기를 이용한 객체인식 기반의 건설 현장 정보 시각화 구현

Augmented Reality Framework to Visualize Information about Construction Resources Based on Object Detection

Pham Hung¹⁾, Nguyen Linh²⁾, 이용주³⁾, 박만우⁴⁾, 송은석⁵⁾

Pham, Hung¹⁾ · Nguyen, Linh²⁾ · Lee, Yong-Ju³⁾ · Park, Man-Woo⁴⁾ · Song, Eun-Seok⁵⁾

Received August 24, 2021; Received September 14, 2021 / Accepted September 14, 2021

ABSTRACT: The augmented reality (AR) has recently become an attractive technology in construction industry, which can play a critical role in realizing smart construction concepts. The AR has a great potential to help construction workers access digitalized information about design and construction more flexibly and efficiently. Though several AR applications have been introduced for on-site made to enhance on-site and off-site tasks, few are utilized in actual construction fields. This paper proposes a new AR framework that provides on-site managers with an opportunity to easily access the information about construction resources such as workers and equipment. The framework records videos with the camera installed on a wearable AR device and streams the video in a server equipped with high-performance processors, which runs an object detection algorithm on the streamed video in real time. The detection results are sent back to the AR device so that menu buttons are visualized on the detected objects in the user's view. A user is allowed to access the information about a worker or equipment appeared in one's view, by touching the menu button visualized on the resource. This paper details implementing parts of the framework, which requires the data transmission between the AR device and the server. It also discusses thoroughly about accompanied issues and the feasibility of the proposed framework.

KEYWORDS: AR(Augmented Reality), Construction Site, Wearable Device, Object Detection, Network Protocol

키워드: 증강현실, 건설현장, 웨어러블 기기, 객체 인식, 네트워크 프로토콜

1. 서론

정보통신기술의 발달로 인해 인공지능, 사물인터넷, 로봇, 가상 현실 등의 다양한 기술들이 여러 분야에서 활용되고 있다. 건설 산업도 마찬가지로 이런 흐름에 발맞춰 무인항공기, 지능형 건설 장비, 가상 현실, 증강 현실, 빅데이터, 인공지능 등의 다양한 기술을 활용하여 생산성 및 품질 향상, 현장 안전관리 개선 등을 도모하고 있다. 이 중 증강 현실(Augmented Reality, AR) 기술은 가상 세계의 데이터와 상호작용이 가능한 환경을 제공하는 기술로서, 최근 여러 분야에서 주목받고 있다. AR은 현실 세계에는 영향을 주지 않고, 자유롭게 가상의 데이터를 이용하거나 가공하는 등의 상호작용이 가능한 환경을 제공함으로써 정보 접

근성과 활용도를 높여주는 정보 시각화 기술이라 할 수 있다.

AR 기술은 사용자가 처한 현실 상황을 우선적으로 인식하고, 이를 바탕으로 사용자에게 필요한 정보를 사용자가 직관적으로 이해할 수 있도록 시각화하는 과정을 거친다. 카메라, GNSS(Global Navigation Satellite System) 등 다양한 센서가 장착된 스마트폰 사용이 보편화되면서 사용자의 시각, 위치, 움직임 등 현실 상황에 대한 인식이 가능해졌고 이로 인해 AR 기술 개발이 다양한 분야에서 활발하게 이뤄져 왔다.

카메라 영상을 통해 집안의 벽과 바닥을 인식하여 가상의 3D 가구 모델을 미리 배치해볼 수 있게 되었고, 옷을 구매하기 전에 본인에게 어울리는지 가상의 옷을 덧씌워 살펴볼 수도 있게 되었다. 이외에도 얼굴을 인식해 가상으로 화장을 해보거나 성별,

¹⁾학생회원, 명지대학교 토목환경공학과 석사과정 (phamhung0306@gmail.com)

²⁾학생회원, 명지대학교 토목환경공학과 석사과정 (linngn@mju.ac.kr)

³⁾학생회원, 명지대학교 토목환경공학과 박사과정 (leetoday@mju.ac.kr) (교신저자)

⁴⁾정회원, 명지대학교 토목환경공학과 부교수 (mwpark@mju.ac.kr)

⁵⁾정회원, 한국도로공사 스마트건설사업단 대리 (ses@ex.co.kr)

나이 등의 바뀌보는 등 엔터테인먼트 영역을 포함하여 일상 영역에서 다양한 형태로 AR 기술이 활용되고 있다. 산업 분야에서도 AR 기술의 각종 활용 사례를 쉽게 찾아볼 수 있다. 복잡한 기계 설비의 작동, 유지관리 방법 등의 내용을 기존에는 문서를 통해 접해야 했지만, AR 기술을 통해 기계 설비 위에 3D 모델을 투영시켜 보여주거나 상황을 인식해 적합한 정보를 작업자의 시야에 표시해주는 등의 형태로 작업자가 장비, 설비 등에 대해 더 쉽게 이해하도록 도와줄 수 있다. AR 기술을 적용한 협업 시스템도 활용되고 있다. 제품의 실물 생산에 앞서 의사결정 과정에서 가상의 제품을 현실 세계에 투영하여 보여줌으로써 의사결정자들이 제품의 실제 모습을 더 직관적으로 이해할 수 있다. 여기에 인터넷 통신 기술을 결합하면 당사자들이 서로 먼 거리에 떨어져 있어도 바로 앞에 있는 것처럼 의사소통할 수도 있다. 공상 과학 영화 등에 등장하는 홀로그램 통화와 비슷한 형태로 볼 수 있다. AR 기술은 이러한 기능들을 통해 사용자에게 더 생생하고 직접적인 경험이 가능한 환경을 제공한다.

건설 산업에서도 비슷한 형태로 설계, 시공, 유지관리 단계 전반에 걸쳐 활용도가 검토되고 있다. 설계 3D 모델을 현장에서 카메라 화면에 중첩시켜 시각화하여 설계 정보를 보다 직관적으로 확인할 수 있도록 도와주는 목적으로의 기술, 설계 도면을 자동으로 인식하여 연관된 3D 모델을 시각화하여 도면 해석을 돕는 목적으로의 기술 등이 시도된 바 있다. 또한, 이미 시공되었으나 직접 눈으로 확인하기 어려운 배관, 전기 설비, 철근, 지하 구조물 등의 3D 모델을 투영하여 시각화하는 기술들도 개발되고 있다. 하지만, 실제 현장에서 활용되고 있는 사례는 찾아보기 어려운 상황이다.

이 연구에서는 기존에 시도되어온 것들과는 다소 다른 목적으로의 건설현장용 AR 기술 체계(Lee et al., 2020)를 소개하고, 그 중 첫 단계의 기술 구현 과정을 검토하고자 한다. Lee et al.(2020)은 웨어러블 AR 기기를 착용한 사용자가 현장에서 장비나 인력에 대한 정보를 확인할 수 있는 AR 기술을 소개한 바 있다. 웨어러블 AR 기기에 장착된 카메라의 영상에 CNN(Convolutional Neural Network) 기반 객체 인식 알고리즘을 적용하여 장비 또는 인력을 인식하고, 인식된 객체에 대한 정보를 디지털 트윈(Digital twin) 모델로부터 전송받아 시각화하는 체계이다. CNN 기반 객체 인식 알고리즘을 실시간으로 연산하기에는 웨어러블 AR 기기의 성능이 부족하기 때문에 PC에서의 알고리즘 실행이 불가피하고, 이를 위해서는 로컬 네트워크를 통한 웨어러블 AR 기기와 PC 간의 데이터 전송 과정이 필요하다. 이 연구에서는 HoloLens 2로 촬영된 영상 데이터에서의 객체 인식 알고리즘 성능을 간단히 검토하고, 이 알고리즘의 실시간 처리를 위해 HoloLens 2와 PC 간의 로컬 네트워킹 과정에 대하여 검토하고자 한다.

2. 기존 연구

2.1 건설 산업과 AR

AR 기술은 컴퓨터를 통해 생성된 가상의 정보를 현실 세계에 투영하여 실시간으로 보여주는 기술이다(Elsbach & Kramer, 2015). AR 기술을 통해 현실 세계와 가상 세계를 잘 조화시킴으로써 사용자가 가상 환경을 실제 환경과 같이 느끼게 하여 사용자와 가상 세계 간의 자연스러운 실시간 상호작용이 가능하도록 몰입도가 높은 환경을 제공해주는 것이 그 역할이다. 다양한 분야에서 활용될 수 있으며 건설 산업에서도 많은 AR 관련 연구가 진행되었다(Behringer et al., 1999; Jiao et al., 2013; Zollmann et al., 2014).

아직까진 초기 단계이지만 몇몇 상용 솔루션들이 이미 시장에 출시되어 있기도 하다. SYNCHRO XR은 HoloLens 2에서 실행 가능한 건설 관련 AR 애플리케이션 중 하나이다(Christine, 2019). 프로젝트의 3D 도면과 공정 등을 확인할 수 있고, 다른 사용자들과의 협업을 위한 커뮤니케이션 기능들을 담고 있다. SYNCHRO XR을 통해 복잡한 프로젝트 관련 정보들을 보기 쉽게 시각화하여 작업 능력을 올리고, 발생 가능한 오류를 줄일 수 있다고 한다. 건설현장보다는 주로 실내에서 프로젝트 관리자들이 사용하기에 적합하도록 제작되었다. 또 다른 애플리케이션인 vGIS는 AR 기반의 BIM, GIS 솔루션으로 현장 사용 목적으로 제작되었다(vGis, 2021). 기존의 BIM 도면 및 GIS 데이터를 실제 작업 현장에서 AR 형태로 열람할 수 있도록 자동으로 전환해준다. 지상 구조물뿐만 아니라 가스, 전기, 상하수도 등의 지하 인프라 시설들까지 현장에서 육안으로는 확인하기 어려운 부분들까지 접근할 수 있다.

2.2 CNN 기반 객체 인식 모델

AR 기기의 카메라를 통해 객체를 인식하기 위해서 영상 기반 객체 인식 알고리즘이 활용된다. 영상 데이터로부터 특정 객체를 인식하기 위한 수많은 컴퓨터 비전 알고리즘 관련 연구가 진행되었다. 객체 인식 알고리즘은 크게 딥러닝 이전과 이후로 나눌 수 있다. 딥러닝 기반의 객체 인식 알고리즘은 기존의 이미지 프로세싱 방식의 알고리즘들과 비교하여 속도가 빠르면서도 정확도 또한 높아서 최근에는 딥러닝 기반의 객체 인식 알고리즘이 많이 활용되고 있다. 본 연구에서는 이들 중에서도 CNN 기반의 객체 인식 알고리즘을 활용한다. CNN 기반 객체 인식 알고리즘으로는 R-CNN(Girshick et al., 2014), Faster R-CNN(Ren et al., 2016), SSD(Liu et al., 2016), YOLO(Redmon et al., 2016) 등이 있다.

객체 인식 기술, 특히 CNN에 기반한 객체 인식 알고리즘들은 기존의 영상 처리 기반의 알고리즘들과 비교해 빠른 속도와 높은

정확도로 인해 여러 산업 분야에서 활용되기 시작했다. 그중에서 YOLO 알고리즘은 다른 알고리즘들과 비교하여 실시간 처리가 가능할 정도로 빠른 처리 속도를 가졌으면서도 정확도가 높다는 것이 특징이다(Kim et al., 2020). YOLO 알고리즘은 최근 출시된 YOLOv4까지 꾸준한 성능 개선이 이루어지고 있다(Jiang et al., 2020). YOLOv4는 처리 속도를 끌어올려 최대 54 fps(frames per second)의 속도로 연산 처리가 가능하다(Bochkovskiy et al., 2020).

건설 산업에서도 객체 인식 알고리즘들을 활용하기 위한 다양한 연구들이 진행되었다. 주로 특정 공정의 생산성 확인을 위해 객체 인식을 활용하거나(Luo et al., 2018; Kim et al., 2018; Fang et al., 2018b), 안전관리를 위해 활용되기도 한다(Fang et al., 2018a; Fang et al., 2018c; Yang et al., 2019). 최근에는 건설현장에서 사용하기 위한 무인기 등의 자율 이동 개발 목적으로 활용되기도 한다(Asadi et al., 2019).

본 연구에서는 웨어러블 AR 기기의 카메라를 통해 건설현장 객체(구조물, 장비, 인력)를 인식하기 위해 CNN 기반의 객체 인식 알고리즘 중 하나인 YOLO 알고리즘을 적용했다.

2.3 웨어러블 AR 기기

AR 시장이 확대되고 활용 영역이 넓어지면서 스마트폰, 태블릿뿐만 아니라 다양한 형태의 AR 기기들이 개발됐다. 그중에서도 Microsoft, Google, Magic Leap, Thirdeyes 등의 제조사에서는 웨어러블 형태의 AR 기기들을 출시하기도 했다. 일반적으로 웨어러블 기기라 하면 사람이 착용하고 자유롭게 움직일 수 있는 형태의 전자기기를 뜻한다. 그중에서도 웨어러블 AR 기기들은 그 특성상 주로 머리에 장착해 AR 콘텐츠를 실행할 수 있는 환경이 함께 제공되는 기기들이다. 가장 대표적인 제품이 Google의 Glass 시리즈와 Microsoft의 HoloLens 시리즈다.

Google Glass의 가장 최신 기종은 Glass Enterprise Edition 2이다(<https://google.com/glass/tech-specs/>). 이 제품은 일반 안경 정도로 크기가 작고 무게 또한 46 g으로 착용 시 불편함이 없다는 것이 가장 큰 장점이다. 그뿐만 아니라 Google의 제품인 만큼 오픈소스 운영체제인 Android OS를 기본으로 지원하여 소프트웨어 개발 환경에 있어 다른 운영체제보다 넓은 저변을 마련하고 있다. 하지만 크기가 작은 만큼 디스플레이, 프로세서, 메모리 등 모든 부분에서 다른 기기들보다 성능이 떨어진다. 특히 디스플레이는 640x360 해상도의 단안 디스플레이를 채택하여 한번에 많은 정보를 표현하기에는 무리가 있고, 사용자가 체감하는 디스플레이의 크기 또한 매우 작아서 정확한 정보 전달이 어려울 수 있다.

Microsoft HoloLens의 가장 최신 기종은 HoloLens 2이다(<https://www.microsoft.com/ko-kr/holoLens/hardware>).

Google의 제품과는 완전히 상반되는 컨셉의 제품으로 Google의 제품과 비교하여 크기가 크고, 무게 또한 566 g으로 10배 이상 차이 난다. 하지만 그만큼 디스플레이, 프로세서, 메모리 등의 성능이 전반적으로 뛰어나 더 다양하고 적극적인 활용이 가능하다. 운영체제는 Windows 10 Holographic OS라는 자체 개발된 OS를 사용한다. HoloLens 시리즈의 가장 큰 특징은 디스플레이이다. 다른 기기들의 일반적인 평면 디스플레이와 달리, 2048x1080 해상도의 양안 홀로그래프 디스플레이를 채택하여 더 사실적이고 선명한 3D 표현이 가능해 AR 기술 활용에 가장 유리하다. 이런 점 때문에 위에서 소개한 SYNCHRO XR, vGIS 등의 건설 AR 솔루션 등에서도 활용되고 있다.

본 연구에서는 객체를 인식하고, 해당 객체의 정보를 시각화하는 것이 목적이기 때문에 기기 운용의 편의성보다는 기기 자체의 성능이 중요하다고 판단하여 Microsoft의 HoloLens 2를 활용했다.

3. 연구 목적 및 범위

건설 프로젝트의 의사결정 과정에 있어 건설현장 정보에 대한 접근성은 매우 중요한 요소이다. 다만 대부분의 건설현장 정보들은 문서로 저장되기 때문에 정보에 대한 접근 편의성이 떨어지고, 직관적인 이해가 어려운 경우가 많다. 중요한 정보들이나 직관적으로 이해하기 어려운 정보들을 웨어러블 AR 기기를 활용하여 표현할 수 있다면 프로젝트 관리자, 작업자들의 작업 이해도를 향상시키는데 많은 도움이 될 것이다. 이런 목적으로 디지털 트윈의 정보들에 접근할 수 있도록 웨어러블 AR 기기를 활용하는 방안에 대한 연구가 진행된 바 있다(Lee et al., 2020). 해당 연구에서는 웨어러블 AR 기기를 활용하기 위한 몇 가지 단계에 대해 설명하고 있으며 각 단계별로 필수적인 요소들이 존재한다(Figure 1).

그 첫 번째는 객체 인식이다. 사용자가 현실 세계에 존재하는 특정 객체에 대한 정보를 웨어러블 AR 기기를 통해 열람하고자 할 때, 웨어러블 AR 기기가 해당 객체를 인식할 수 없다면 어떤 객체의 정보를 불러와야 할지 결정할 수 없기 때문에 어떤 방식으로든 AR 기기가 객체를 인식할 수 있도록 하는 과정이 필요하다. 본 연구에서는 해당 과정에서 카메라를 통한 영상 기반 객체 인식 기술을 적용하고자 한다. 이 과정에서 검증된 몇 가지의 CNN 기반 객체 인식 알고리즘들이 고려됐으며 그 중 YOLOv3와 YOLOv4의 비교를 통해 적합한 객체 인식 알고리즘을 선정했다.

두 번째는 기기 간의 통신이다. 디지털 트윈은 건설현장의 수많은 정보를 담고 있으며 수시로 업데이트되는 일종의 네트워크 서버 역할을 한다. 디지털 트윈의 정보를 열람하기 위해서는 기본적으로 특정 정보를 열람하기 위한 요청 신호를 전송할 수 있

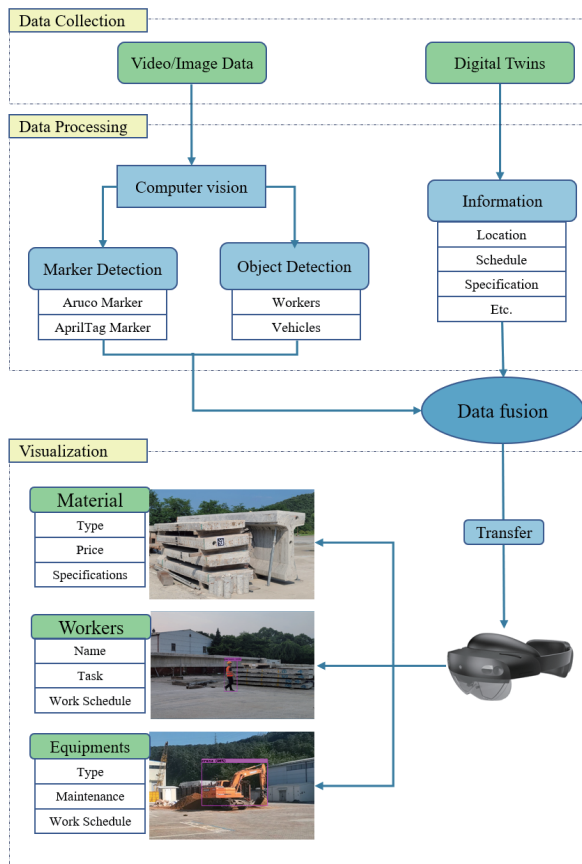


Figure 1. AR framework for visualizing the information of on-site resources (materials, worker, and equipment)

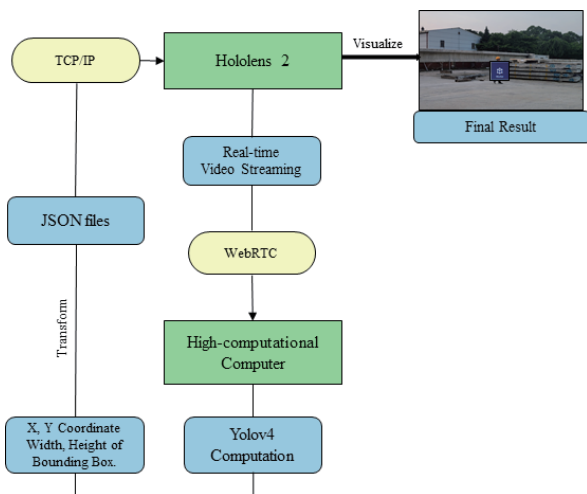


Figure 2. Data communication between Hololens 2 and PC

여야 하고, 당연하게도 해당 데이터를 받아들일 수 있어야 한다. 그 뿐만 아니라 크고 무거울수록 사용성이 저하될 수밖에 없는 웨어러블 AR 기기의 특성상 고성능을 요구하는 연산 처리를 수행하기에는 무리가 있다. 고용량 연산 처리가 필요한 경우에는 클라우드 컴퓨팅을 이용할 수 있도록 하기 위해서도 기기 간의 통신

은 반드시 필요하다. 본 연구에서는 이 과정을 로컬 네트워크로 진행했으며 데이터의 입력, 출력 간에 발생할 수 있는 지연 시간을 최소화하기 위한 통신 프로토콜에 대해 확인하고, 그 사용성을 확인한다.

마지막은 시각화 과정이다. 3차원 홀로그래픽 디스플레이를 사용하는 Hololens 2의 특성상 디스플레이상에 어떤 내용을 표시하기 위해서는 반드시 3차원 좌표가 필요하다. 하지만 영상 기반 객체 인식 과정에서 카메라를 통해 데이터를 취득하기 때문에 거리 정보가 손실되어 인식된 객체의 정확한 3차원 위치를 획득할 수 없게 된다. 이로 인해 3차원 디스플레이상에 해당 객체를 표시하기 위해서는 별도의 과정을 거쳐야 할 필요가 있다. 본 연구에서는 이 방법에 대해 제안하고, 그 사용성을 확인한다.

4. 연구 방법

여기서는 앞에서 소개된 AR 기술의 구성 단계들 중 이 연구 범위에 해당하는 부분들을 세부적으로 다루고자 한다. Figure 1의 내용 중 포함되지 않는 부분은 Hololens 2의 영상에서 인식된 객체들과 디지털 트윈에 저장되어 있는 객체별 정보들을 연계하는 부분이다. 이 연구에서 다루는 부분의 프레임워크는 Figure 2에 간략하게 나타냈으며 크게 세 과정으로 나뉘어 수 있다. 첫 번째로, Hololens 2에서 촬영된 영상을 네트워크를 통해 고성능 컴퓨터로 전송한다. 두 번째로, 전송 받은 영상 데이터를 고성능 컴퓨터에서 연산하여 건설 인력 또는 장비 등을 인식한다. 마지막으로 객체 인식 결과를 Hololens 2로 전송하고, 사용자가 객체에 대한 정보들을 필요 시 시각화할 수 있도록 해당 객체 위에 메뉴 버튼을 시각화한다. 첫 번째 과정에는 실시간 영상 스트리밍에 적합한 프로토콜인 WebRTC(Web Real-Time Communication)이 사용되었고, 두 번째 과정에는 객체 인식 모델 중 하나인 YOLOv4가 선정되어 사용되었으며, 마지막 과정에서는 TCP/IP 연결을 통해 JSON(JavaScript Object Notation) 형식으로 데이터를 전송하는 방식이 사용되었다. 다음 세부 절에서 각 단계별로 사용된 방식이나 알고리즘들에 대하여 설명하고자 한다.

4.1 Hololens 2로부터의 영상 스트리밍

Hololens 2에는 800만 화소의 카메라가 장착되어 있으며 최대 해상도 1920x1080, 30 fps의 동영상 촬영이 가능하다. 프로세서로는 Qualcomm Snapdragon 850이 탑재되어 준수한 계산 성능을 보이지만 해상도 1920x1080 영상에 YOLOv4와 같은 CNN 기반의 객체 인식 알고리즘을 적용하여 실시간으로 처리하기에는 부족하다. 이러한 이유로 Hololens 2에서 직접적인 객체 인식 과정을 처리하지 못하고 별도의 고성능 컴퓨터에서 해당 과

정을 대신하는 것이다. 따라서, Hololens 2에서 고성능 컴퓨터로 영상을 스트리밍할 수 있어야 한다. 이 과정에서 영상 정보의 손실과 스트리밍 지연 시간을 최소화할 필요가 있다. 본 연구에서는 Mixed Reality Streaming과 WebRTC, 두 가지의 영상 스트리밍 프로토콜을 적용하여 비교해보았다.

Mixed Reality Streaming은 Microsoft에서 제공하는 소프트웨어로, Hololens 2에서 촬영된 영상을 파편화된 여러 개의 파일로 나누어 저장하고, 이를 네트워크를 통해 전송, 재결합하는 방식의 영상 전송 프로토콜이다. WebRTC는 실시간 데이터 스트리밍을 위해 만들어진 오픈 소스 소프트웨어이다. P2P(Peer-to-Peer) 데이터 송수신, 음성/영상 통화 등의 기능을 별도의 소프트웨어나 플러그인 없이 구현할 수 있도록 해준다(Sredojev et al., 2015). 두 가지 프로토콜 모두 기본적으로 동일한 기능을 수행하며 Hololens 2에서의 실시간 영상 스트리밍 목적으로 활용이 가능하다. 두 프로토콜의 가장 중요한 차이점은 데이터 전송 및 취합 방식에 있다. 두 프로토콜 중에서 WebRTC가 이 연구 목적에 더 적합하다고 판단하여 WebRTC를 적용했다. 이것에 대한 내용은 뒤에서 더 자세하게 다룬다.

4.2 객체 인식

두번째 과정은 CNN 기반의 객체 인식 모델을 적용하여 Hololens 2로부터 WebRTC를 통해 전송받은 영상을 처리해, 특정 객체를 인식하는 과정이다. 본 연구의 목적에 맞도록 이 과정은 실시간 혹은 그에 준하는 빠른 연산 속도로 처리 가능해야 한다. 이러한 이유로 본 연구에서는 YOLOv4를 객체 인식 모델로 채택했으며 특정 데이터 세트를 활용해 학습시켰다. 학습된 객체 인식 모델은 굴삭기, 덤프트럭, 레미콘과 같은 건설 장비 및 작업자를 구분할 수 있다.

YOLOv4는 CNN 기반의 여러 객체 인식 모델들과 비교하여 빠른 추론 속도와 높은 인식률을 가진 편에 속한다(Jiang et al., 2020). 이 연구에서는 인식 대상을 건설 인력과 장비로 제한하여 검토하였다. 건설 인력은 일반적인 사람인식으로 대체할 수 있었고, 이 과정에는 특별한 별도의 학습과정을 거치지 않고 기존에 MS COCO(Microsoft Common Objects in Context) 데이터 세트로 학습된 네트워크(Lin et al., 2014)를 그대로 사용하였다. 사람과 달리 건설장비는 기존에 학습되어 공유되고 있는 네트워크 모델이 없기 때문에, 별도의 학습 과정이 필요하였다. 건설장비 모델의 학습을 위해 ACID(Alberta Construction Image Dataset) 데이터 세트(Xiao & Kang, 2021)를 사용하였다. ACID 데이터 세트는 굴삭기, 덤프트럭, 레미콘의 세 가지의 건설 장비 사진들을 포함하고 있으며 총 2,850 장의 사진 데이터로 구성되어 있다. 이 중 1,995 장을 학습 세트(Training set)로 사용했으며, 이 중 10%를 검증 세트(Validation set)로 사용했다. 나머지 855 장의 이미

지는 테스트 세트(Test set)으로 사용하여 성능을 검토하였다.

4.3 객체 인식 결과 전송 및 시각화

각각의 영상 프레임에 대해 객체 인식 알고리즘이 적용된 결과로 프레임에서 객체가 차지하는 영역을 저장하는데, 일반적으로 객체를 포함하는 사각형 영역으로 표현한다. 이 사각형 영역에 대한 정보를 전달하기 위해서는 최소한 사각형의 중심 좌표(x_c, y_c), 폭, 높이 등 네 가지 데이터가 저장되어야 한다. 이 사각형 정보들은 다시 Hololens 2로 전송하기 위하여 JSON 형식으로 저장된다. JSON 형식은 데이터가 가벼워 네트워크 전송 시간과 컴퓨터 처리 속도에 있어 유리하다. 저장된 JSON 형식의 파일은 TCP/IP를 통해 Hololens 2로 전송되는데 이는 별도의 어플리케이션에서 구동된다. 어플리케이션은 Unity 엔진과 MRTK(Mixed Reality ToolKit)(Microsoft, 2019) 도구들을 활용하여 Hololens 2 환경에서 작동할 수 있도록 제작하였다. 결과 데이터는 Unity 엔진의 simpleJSON(Ippolito, 2018) 라이브러리를 통해 문자열 데이터로 전환해 사용한다.

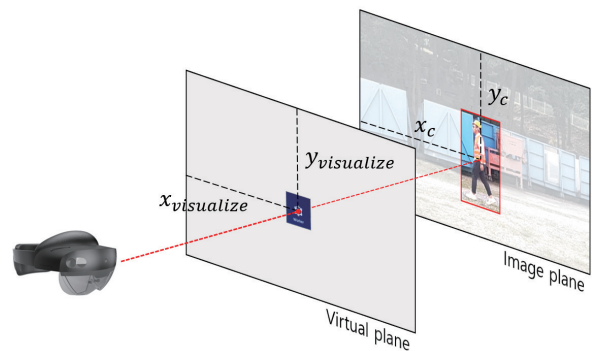


Figure 3. Localization of a menu button in 3D

다음 과정은 인식된 객체 위에 메뉴 버튼을 시각화하는 것이다. Hololens 2에서는 모든 시각화가 3D로 이뤄지기 때문에 메뉴 버튼의 위치 역시 3D 좌표로 지정해줘야 하는 반면에, 전송된 객체 인식 결과는 영상 프레임에서의 2D 좌표이기 때문에 3D 좌표로의 변환이 필요하다. 본 연구에서는 메뉴 버튼의 3D 좌표 설정을 위해 Hololens 2 사용자 시점에 직각인 가상 평면을 만들어서 활용하였다. 사용자 시점을 기준으로 객체 인식 결과인 2D 좌표로 이어지는 벡터를 가상 평면에 투영하여 해당 위치에 버튼 형태의 3D 객체를 시각화하도록 하였다. 가상 평면은 Hololens 2의 디스플레이와 같은 가로, 세로 3:2의 비율로 설정했다. 가상 평면은 Hololens 2로부터 사용자가 어느 방향을 바라보든 고정된 거리, 각도에 위치한다. 가상 평면 자체는 투명한 상태를 유지하여 사용자가 현실 세계를 바라볼 때 시야를 방해하지 않는다. 가상 평면은 (10, 10)의 임의의 좌표계로 나누었으며 영상의 좌표계인 (1080, 720)을 아래와 같이 변환한다(Figure 3).

$$x_{visualize} = x_c \times \frac{10}{1080}$$

$$y_{visualize} = y_c \times \frac{10}{720}$$

가상 평면에 인식 결과의 버튼을 표시하면 Hololens 2를 통해 바라보는 사용자의 시선에서는 피사체(인식 목표 객체)에 버튼이 겹쳐진 형태로 자연스럽게 표현된다(Figure 4). 가상 평면에 표현되는 버튼은 YOLOv4의 인식 결과와 마찬가지로 매 프레임마다 업데이트되기 때문에 피사체의 위치와 거의 실시간으로 일치하게 된다.

5. 실험 및 결과

본 연구의 최종 결과는 앞서 언급한 일련의 과정을 거쳐서 이루어지므로 과정 간의 충돌이나 성능 변화 없이 통합될 수 있도록 각 과정을 개별적으로 실험하고, 검증했다. 최종적으로, 여러 방법 중 높은 정확도와 실시간 사용이 가능할 정도의 연산 속도를 확보할 수 있는 최적의 방법을 선정했다.

Table 1. Time delay caused when using mixed reality streaming or WebRTC

Protocol	fps	Time delay (second)
Mixed Reality Streaming	82	2.73
WebRTC	6	0.2

5.1 Hololens 2로부터의 영상 스트리밍 테스트

Hololens 2를 활용해 영상을 스트리밍하면서 Mixed Reality Streaming과 WebRTC를 모두 적용하고 비교했다. 두 프로토콜은 모두 영상 스트리밍 목적으로 사용할 수 있으며 Hololens 2와의 연결성도 유사하지만, 영상 전송 지연 시간 측면에서는 현격한 차이를 보였다. Table 1은 WebRTC와 Mixed Reality Streaming를 사용했을 때의 초당 전송 프레임 수와 영상 전송 지연 시간을 보이고 있다. 여기서 지연 시간은 Hololens 2에서 영상 프레임이 투영된 시점으로부터 객체 인식 결과가 표시되는 시점까지 소요된 시간을 측정하는 것이다. Hololens 2의 카메라는 최대 1920x1080의 해상도를 지원하지만 객체 인식 과정의 연산 부하량과 네트워크 전송 대역폭 등을 고려하여 1080x720으로 제한했다. 이를 통해, 일정 이상의 객체 인식 성능을 유지하면서 30 fps의 일관된 속도로 안정적인 데이터 전송이 가능하다.

Mixed Reality Streaming을 이용했을 때의 지연 시간은 2.73 초, WebRTC를 이용했을 때의 지연 시간은 0.2 초로 지연 시간이 크게 단축된 것을 확인할 수 있다. 이 차이는 각 프로토콜의 처리 방식 차이로 인해 발생한다. Mixed Reality Streaming은 캡



Figure 4. Visualization of a button on each detected worker in the Hololens 2 user's view

처된 영상을 파편화된 mp4 형식으로 저장하고, 네트워크를 통해 전송한 뒤, 다시 연속적인 영상 형태로 통합하는 과정을 거친다. 반면, WebRTC 프로토콜은 영상 분할, 재결합 과정 없이 영상 스트리밍을 지속적으로 전송한다. 결과적으로 지연 시간은 인코딩, 전송 및 디코딩 과정에서만 발생하기 때문에 상대적으로 빠른 지연 시간을 확보할 수 있다. 따라서, Hololens 2에서 연산용 컴퓨터로 영상을 전송하는데 있어 WebRTC 프로토콜이 적합하다.

5.2 객체 인식

객체 인식 성능 검토를 위해 Hololens 2를 이용하여 촬영된 해상도 1080x720의 영상에 연산속도가 비교적 빠른 YOLO 알고리즘을 적용하여 테스트하였다. YOLO 알고리즘들 중 YOLOv3와 YOLOv4의 두 가지 버전을 적용하여 성능을 비교해보았다. 사람 인식 네트워크는 이미 기존에 MS COCO 데이터 세트를 이용해 학습된 네트워크를 그대로 사용하였기에 별도의 성능 검토를 하지 않았으며, ACID 데이터 세트로 새롭게 학습한 건설 장비 인식 네트워크의 성능만 비교하여 Table 2에 정리하였다.

Table 2에서 YOLOv4가 YOLOv3보다 인식 성능이 확연하게 우수한 것을 볼 수 있다. 알고리즘 연산은 고성능 그래픽 카드가 장착된 일반 PC에서 이뤄졌으며, YOLOv4 적용 시 30 fps 영상을 실시간으로 처리하는데에 무리가 없음을 확인할 수 있었

Table 2. Performance of YOLOv3 and YOLOv4 trained with the ACID dataset

	Class name	Average Precision (%)	mAP _{0.5}
YOLOv3	Excavator	81.83	75.99
	Dump Truck	64.12	
	Concrete Mixer Truck	82.02	
YOLOv4	Excavator	97.47	95.42
	Dump Truck	89.87	
	Concrete Mixer Truck	99	

Table 3. Displacement of the indicator depending on the distance of a worker from the Hololens 2 user

Distance (m)	Experiment Number	Displacement (pixel)			
		Minimum	Maximum	Mean	Standard Deviation
5	1	47	70	56.5	8.017
	2	53	69	63.4	5.739
	3	49	60	54.2	4.104
10	1	28	43	34.4	5.016
	2	34	45	39.1	3.510
	3	29	43	36.4	5.542
15	1	13	31	21.5	6.223
	2	13	27	22.1	4.332
	3	13	30	21.7	5.658
20	1	6	20	11.6	4.142
	2	6	15	10.2	3.327
	3	6	14	9.7	2.908

Table 4. Displacement of the indicator depending on the walking speed of a worker

Speed (m/s)	Experiment Number	Displacement (pixel)			
		Minimum	Maximum	Mean	Standard Deviation
0.9	1	22	31	26.8	3.084
	2	10	17	13.2	2.299
	3	19	36	25.2	5.073
1.2	1	18	30	24.3	4.218
	2	24	36	27.8	3.706
	3	10	20	14.4	3.596
1.6	1	33	44	39	3.972
	2	16	30	23.6	5.125
	3	28	34	31.1	2.470

다. 다만, 일부분이 가려진 장비들의 인식은 제대로 이뤄지지 않았다. 즉, 장비 전체가 모두 영상에 온전하게 보일 때에만 인식이 잘 이뤄졌다. 이는 학습 데이터 세트의 한계로 보인다. 학습 데이터의 사진들에도 일부분 가려진 장비들이 포함되어 있으나, 레이블링(labeling)이 되어 있지 않은 것으로 확인되었다. 추후, 가려진 상태의 장비들에 대한 레이블링이 추가적으로 수행되고, 지속적으로 학습 데이터들을 축적해나감에 따라 점차적으로 해결할 수 있는 문제로 판단된다.

6만 장 이상의 사람 사진을 담고 있는 MS COCO 데이터 세트에는 다양한 자세의 사람들, 다양한 옷을 입은 사람들이 포함되어 있고, 팔, 다리, 머리 등 일부만 레이블링된 사진들도 포함되어 있다. 따라서, 다양한 자세의 인력들, 안전모나 안전 조끼를 착용한 인력들을 정확하게 인식하고, 화면에서 일부가 가려져도 쉽게 인식하는 것을 확인할 수 있었다.

5.3 객체 인식 결과 전송 및 시각화 테스트

Hololens 2에서 객체 인식 결과를 시각화하기 위해서 Unity 2020.3.1.f1 LTS 엔진과 MRTK를 활용해 어플리케이션을 개발했다. 해당 어플리케이션은 연산용 컴퓨터로부터 인식 결과를 전송 받아 사용자의 Hololens 2 디스플레이 상에 결과를 표시한다. 이 과정은 평균적으로 약 160 ms의 시간이 소요된다.

Hololens 2와 객체 인식 모델 간의 상호운용성에 대한 후속 연구 진행에 용이하도록 객체 인식 결과를 사각형 상자 대신 사용자가 상호작용이 가능한 버튼으로 대체했다(Figure 4). 버튼을 통해 사용자가 상호작용이 가능한 메뉴를 불러오거나 정보를 표시하도록 할 수 있다. 해당 버튼의 중심점이 탐지된 객체의 중심점에 위치하도록 설정했다. 본 연구에서 상호운용성에 대한 측정은 AR 기기를 사용하는 사용자의 입장에서 바라볼 때 실제 목표 탐지 객체(중심점)와 객체 인식 모델을 통해 탐지된 객체의 표시(버

튼 중심점) 위치가 얼마나 잘 일치하는지를 pixel 단위로 계산하여 확인했다. 위에서 언급했듯이 사용자가 최종적인 결과를 확인하기까지 영상 전송, 객체 탐지, 결과 수신, 시각화의 단계를 거치면서 지연 시간이 발생하게 된다. 이에 따라, 목표 탐지 객체가 움직이거나 사용자의 시선이 움직이게 되면 알고리즘 연산 시간으로 인한 지연으로 목표 탐지 객체의 중심점과 버튼의 중심점 사이에 오차가 발생하게 되고, 움직임이 클수록 그 오차 또한 커진다. 해당 오차의 정확한 원인을 분석하기 위해 본 연구에서는 두 가지 경우를 고려하여 실험을 진행했다. (1) 작업자(목표 탐지 객체)가 Hololens 2로부터 몇 가지의 다른 거리에서 일정한 속도로 보행하는 경우, (2) 작업자가 Hololens 2로부터 고정된 거리에서 속도를 변화시키면서 이동하는 경우, Table 3와 4는 해당 실험의 결과이며 오차의 단위는 pixel이다.

첫 번째 실험에서 작업자의 이동속도는 평균적으로 1.2 m/s를 유지하도록 했으며 Hololens 2로부터의 거리는 5 m, 10 m, 15 m, 20 m로 나누어 실험했다(Table 3). 5 m 위치에서 최대 70 pixels의 오차가 발생했으나 20 m 위치에서는 그보다 훨씬 적은 최대 20 pixels의 오차만이 발생했다. 이를 통해 거리가 가까울수록 오차가 커지는 것을 확인할 수 있었다. 객체가 카메라와 가까울수록 영상에서 차지하는 pixel 영역의 크기가 커지고, 영상 내에서의 움직임(pixel 변화) 또한 커지게 되는 것이 원인이다. 두 번째 실험에서 작업자는 Hololens 2로부터 15 m 떨어진 곳에 위치했으며 이동속도는 0.9 m/s에서 1.6 m/s까지 변화시켰다. 결과를 통해 속도가 빨라질수록 오차가 커지는 것을 확인할 수 있었다(Table 4).

위의 두 실험에서 연산 처리 및 시각화까지 전체 과정의 처리 시간은 평균적으로 약 500 ms 정도로 측정됐다. 또한, 탐지 객체가 Hololens 2로부터 10~20 m 사이에 위치하고, 1.2 m/s 정도의 속도로 움직일 때, 본 연구에서 제안하는 방법의 성능이 가장 좋은 것으로 나타났다. 두 실험에서 모두 오차의 표준편차가 상대적으로 적은 것에서 알 수 있듯이 실제 Hololens 2를 착용한 사용자 시선에서는 일관된 사용성을 확인할 수 있었다.

그러나 인식 결과의 표시에 있어서 표현 가능 영역이 Hololens 2의 디스플레이 화각 안으로 제한되기 때문에 해당 화각 밖의 인식 결과에 대해서는 표현이 불가하다는 한계가 있다. Hololens 2가 웨어러블 AR 기기 중에서는 가장 넓은 화각을 가졌다는 점을 고려할 때, 본 연구에서 제안하는 시각화 방법을 더 넓은 화각을 가진 기기에 적용한다면 지금보다도 우수한 결과를 확인할 수 있을 것으로 기대할 수 있다.

6. 결론

스마트 건설에서 건설현장 관리를 위한 모니터링 기술의 중요성이 대두되고 있다. 건설현장의 다양하고 복잡한 데이터를 쉽고 직관적으로 열람할 수 있도록 하는 것은 프로젝트 관리 및 의사결정 과정에 있어 매우 중요한 요소 중 하나이다. 그럼에도 불구하고 아직까지 종이 등의 하드 카피 매체를 활용하는 것이 일반적이며 최근에는 데스크탑, 스마트폰, 태블릿PC 등의 전자기기를 활용하기도 하지만 정보 열람 과정이 복잡하다. 이런 방법들

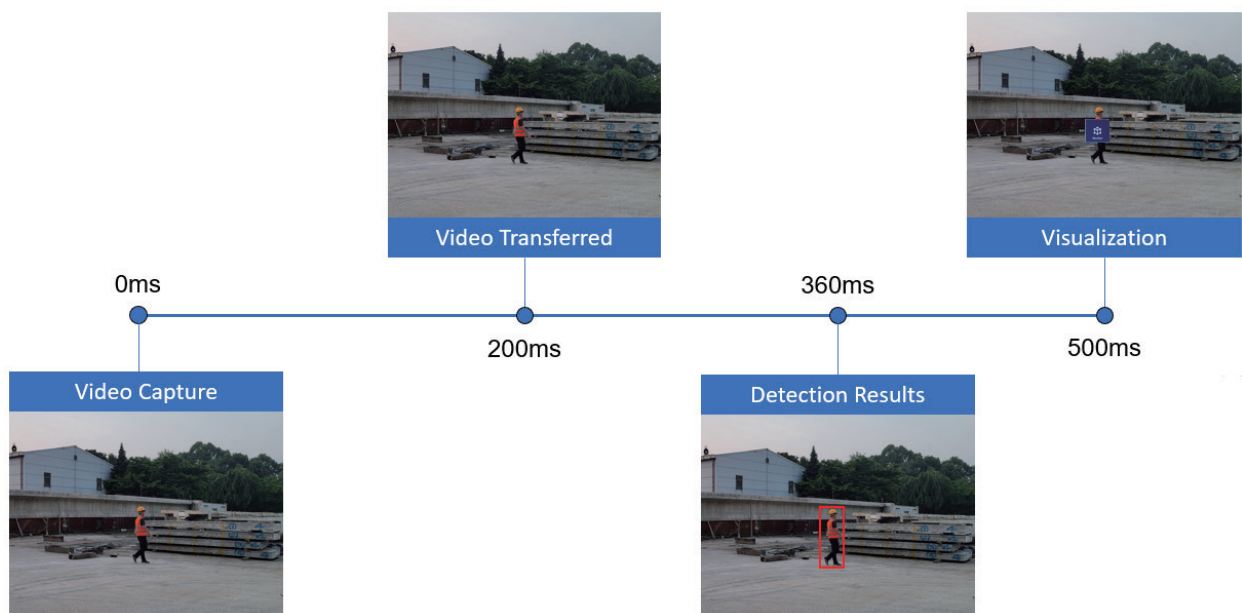


Figure 5. Required processing time from capturing a frame to visualizing a button (500 ms total)

은 접근 가능한 정보가 한정적이고, 현장에서 보유하지 않은 정보에 접근이 필요한 경우 상당한 시간이 소요되기도 하므로 프로젝트 진행에 있어 시간을 지연시키는 원인이 되기도 한다. 최근 웨어러블 AR 기기의 발전으로 인해 건설현장에서 양손을 자유롭게 움직이면서도 정보에 대한 접근이 가능해졌다.

본 연구에서는 웨어러블 AR 기기를 활용하기 위해 CNN 기반 객체 인식 기술을 사용했다. 기존의 프레임워크에 기반하여 실제 하드웨어와 소프트웨어의 상호운용성을 확인하는 차원에서 HoloLens 2와 YOLOv4를 선택했다. HoloLens 2는 YOLOv4 알고리즘을 기반으로 한 객체 인식 모델에 대한 입력 데이터를 수집하여 제공하고, 다시 YOLOv4 알고리즘으로부터 검출된 결과 데이터를 제공받아 사용자의 시야에 출력하는 역할을 한다. 본 연구에서는 대상을 건설 작업자 및 건설 장비로 한정했다. 후속 연구의 목적으로 인식 결과는 버튼의 형태로 표현했으며 버튼 중심이 목표 인식 객체의 중심에 위치하도록 했다.

본 연구에서는 크게 세 가지의 과정을 다룬다. 첫 번째는 HoloLens 2로부터 영상 데이터 전송을 위한 네트워크 프로토콜 선정이다. Mixed Reality Streaming과 WebRTC의 차이점을 확인하고, WebRTC를 적용함으로써 데이터 전송 지연 시간을 획기적으로 줄일 수 있었다. 두 번째는 객체 인식 진행을 위한 알고리즘 선정이다. 영상 데이터를 전송받은 연산용 컴퓨터에서 실제 객체 인식 과정을 진행하는 것으로, 여러 CNN 기반 객체 인식 알고리즘 중 YOLOv3와 YOLOv4를 비교했으며 최종적으로 YOLOv4를 채택했다. 이를 통해, 높은 정확도를 가지면서도 30 fps 수준의 실시간으로 적용 가능한 수준의 객체 인식 속도를 확보할 수 있었다. 마지막은 결과 데이터의 시각화 과정으로 2D 데이터로 이루어진 객체 인식 결과를 3D 좌표를 가진 HoloLens 2에서 시각화하는 방법에 대해 제안하고, 그 사용성을 확인했다. 전체 과정에서 발생한 지연 시간은 약 500 ms 정도로, 대부분의 지연 시간이 장치 간의 네트워크 통신에서 발생했다(Figure 5). 만약 HoloLens 2와 같은 웨어러블 AR 기기의 프로세서가 충분한 연산 성능을 확보할 수 있다면 네트워크 통신이 필요하지 않기 때문에 지연 시간을 획기적으로 줄일 수 있을 것이며 이로 인해 웨어러블 AR 기기와 객체 인식 알고리즘 간의 더 나은 상호운용성을 기대할 수 있을 것으로 예상된다.

본 연구의 한계는 실제 건설 현장의 다양한 변수를 반영하지 못한다는 것에 있다. 건설 인력만을 대상으로 실험이 진행됐으므로 건설 장비, 자재 등의 다양한 객체가 알고리즘에 추가된다면 객체 인식 정확도에 영향이 있을 것으로 예상된다. 뿐만 아니라, 영상 기반 객체 인식 알고리즘을 사용하기 때문에 다양한 환경에서의 영상 데이터를 활용하여 사용성을 검토할 필요가 있다. 주변 환경, 객체와의 거리, 촬영 각도 등의 다양한 인자에 대한 검토가 필요할 것으로 예상된다.

감사의 글

본 연구는 국토교통부/국토교통과학기술진흥원의 지원으로 수행되었음(스마트 건설기술 개발사업 : 과제번호 20SMIP-A157351-02)

References

- Asadi, K., Chen, P., Han, K., Wu, T., Lobaton, E. (2019). Real-time Scene Segmentation Using a Light Deep Neural Network Architecture for Autonomous Robot Navigation on Construction Sites, The 2019 ASCE International Conference on Computing in Civil Engineering, arXiv:1901.08630 [cs.RO], <https://arxiv.org/abs/1901.08630>.
- Behringer, R., Klinker, G., Mizell, D. (1999). Augmented Reality: Placing Artificial Objects in Real Scenes, CRC Press.
- Bochkovskiy, A., Wang, C.-Y., Liao, H.-Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection, Computer Vision and Pattern Recognition, arXiv:2004.10934 [cs.CV], <https://arxiv.org/abs/2004.10934>.
- Christine, B. (2019). Bentley's SYNCHRO XR and Microsoft HoloLens 2 bring the benefits of mixed reality to construction sites, <https://www.bentley.com/ko/about-us/news/2019/february/24/synchro-xr-with-holoLens-2-release> (May. 21, 2021).
- Elsbach, K. D., Kramer, R. M. (2015). Handbook of Qualitative Organizational Research: Innovative Pathways and Methods, Routledge.
- Fang, Q., Li, H., Luo, X., Ding, L., Rose, T. M., An, W., Yu, Y. (2018). A deep learning-based method for detecting non-certified work on construction sites, Advanced Engineering Informatics, 35, pp. 56-68.
- Fang, W., Ding, L., Zhong, B., Love, P. E.D., Luo, H. (2018). Automated detection of workers and heavy equipment on construction sites: A convolutional neural network approach, Advanced Engineering Informatics, 37, pp. 139-149.
- Fang, Y., Ding, L., Luo, H., Love, P. E.D. (2018). Falls from heights: A computer vision-based approach for safety harness detection, Automation in Construction, 91, pp. 53-61.

- Girshick, R., Donahue, J., Darrell, T., Malik, J. (2014). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 580–587.
- Ippolito, B. (2018). simplejson — JSON encoder and decoder, <https://simplejson.readthedocs.io/en/latest/> (May. 06, 2021).
- Jiang, Z., Zhao, L., Li, S., Jia, Y. (2020). Real-time object detection method based on improved YOLOv4-tiny, Computer Vision and Pattern Recognition, arXiv:2011.04244 [cs.CV], <https://arxiv.org/abs/2011.04244>.
- Jiao, Y., Zhang, S., Li, Y., Wang, Y., Yang, B. (2013). Towards cloud Augmented Reality for construction application by BIM and SNS integration, Automation in Construction, 33, pp. 37–47.
- Kim, H., Bang, S., Jeong, H., Ham, Y., Kim, H. (2018). Analyzing context and productivity of tunnel earthmoving processes using imaging and simulation, Automation in Construction, 92, pp. 188–198.
- Kim, J., Sung, J.-Y., Park, S. (2020). Comparison of Faster-RCNN, YOLO, and SSD for Real-Time Vehicle Type Recognition, 2020 IEEE International Conference on Consumer Electronics – Asia (ICCE-Asia), 2020, pp. 1–4.
- Lee, Y.-J., Kim, J.-Y., Pham, H., Park, M.-W. (2020). Augmented Reality Framework for Efficient Access to Schedule Information on Construction Sites, Journal of KIBIM, 10(4), pp. 60–69.
- Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., Dollár, P. (2014). Microsoft COCO: Common Objects in Context, ECCV 2014: Computer Vision – ECCV 2014, pp. 740–755.
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., Berg, A. C. (2016). SSD: Single Shot MultiBox Detector, Computer Vision – ECCV 2016, pp. 21–37.
- Luo, H., Xiong, C., Fang, W., Love, P. E.D., Zhang, B., Ouyang, X. (2018). Convolutional neural networks: Computer vision-based workforce activity assessment in construction, Automation in Construction, 94, pp. 282–289.
- Microsoft. (2019). Introducing MRTK for Unity, <https://docs.microsoft.com/ko-kr/windows/mixed-reality/develop/unity/mrtk-getting-started> (May. 04, 2021).
- Redmon, J., Divvala, S., Girshick, S., Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779–788.
- Ren, S., He, K., Girshick, R., Sun, Jian. (2016). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), pp. 1137–1149.
- Sredojević, B., Samardžija, D., Posarac, D. (2015). WebRTC technology overview and signaling solution design and implementation, 38th International Convention on Information and Communication Technology Electronics and Microelectronics (MIPRO), 2015, pp. 1006–1009.
- vGis. (2021). BIM and GIS Data In Augmented Reality, <https://www.vgis.io/esri-augmented-reality-gis-ar-for-utilities-municipalities-locate-and-municipal-service-companies/> (May. 22, 2021).
- Xiao, B., Kang, S.-C. (2021). Development of an Image Data Set of Construction Machines for Deep Learning Object Detection, Journal of Computing in Civil Engineering, 35(2), [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000945](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000945).
- Yang, Z., Yuan, Y., Zhang, M., Zhao, X., Zhang, Y., Tian, B. (2019). Safety Distance Identification for Crane Drivers Based on Mask R-CNN, Sensors, 19(12), <https://doi.org/10.3390/s19122789>.
- Zollmann, S., Hoppe, C., Kluckner, S., Poglitsch, C., Bischof, H., Reitmayr, G. (2014). Augmented Reality for Construction Site Monitoring and Documentation, Proceedings of the IEEE, 102(2), pp. 137–154.