

액티브 러닝을 활용한 영상기반 건설현장 물체 자동 인식 프레임워크

김진우* · 지석호** · 서준오***

Kim, Jinwoo* · Chi, Seokho** · Seo, JoonOh***

Automated Vision-based Construction Object Detection Using Active Learning

ABSTRACT

Over the last decade, many researchers have investigated a number of vision-based construction object detection algorithms for the purpose of construction site monitoring. However, previous methods require the ground truth labeling, which is a process of manually marking types and locations of target objects from training image data, and thus a large amount of time and effort is being wasted. To address this drawback, this paper proposes a vision-based construction object detection framework that employs an active learning technique while reducing manual labeling efforts. For the validation, the research team performed experiments using an open construction benchmark dataset. The results showed that the method was able to successfully detect construction objects that have various visual characteristics, and also indicated that it is possible to develop the high performance of an object detection model using smaller amount of training data and less iterative training steps compared to the previous approaches. The findings of this study can be used to reduce the manual labeling processes and minimize the time and costs required to build a training database.

Key words : Construction object, Vision-based, Object detection, Active learning, Computer vision

초 록

최근 많은 연구자들이 대규모 현장에 투입된 건설자원의 유형과 위치를 자동 파악하는 영상분석기술을 활발히 개발하고 있다. 하지만 기존의 방법들은 인식하고자 하는 건설 물체(작업자, 중장비, 자재 등)를 학습용 이미지 데이터에 표시하는 Labeling 작업을 요구하고 이에 불필요한 시간과 노력이 낭비된다는 한계가 있다. 이러한 한계를 보완하기 위해서 본 연구는 액티브 러닝을 활용한 영상기반 건설현장 물체 자동 인식 프레임워크를 제안함을 목표로 한다. 개발 프레임워크 검증용 목적으로 건설분야 Benchmark 데이터셋을 이용하여 실제 실험을 진행하였다. 그 결과, 액티브 러닝을 통해 학습한 모델은 다양한 특성을 지닌 건설물체를 성공적으로 인식할 수 있었고, 기존의 학습 DB 구축 방식과 비교할 때 적은 데이터 수와 반복학습 횟수로도 높은 성능을 가지는 영상분석모델을 개발할 수 있었다. 결과적으로 기존에 요구되던 학습 DB 구축을 위한 Labeling 작업을 줄일 뿐만 아니라 총 시간과 비용을 최소화할 수 있다.

검색어 : 건설현장, 영상기반, 물체 인식, 액티브 러닝, 컴퓨터 비전

* 정희원 · 서울대학교 건설환경종합연구소 선임연구원 (Seoul National University · jinwoo92@snu.ac.kr)

** 종신회원 · 교신저자 · 서울대학교 건설환경공학부 부교수 (Corresponding Author · Seoul National University · shchi@snu.ac.kr)

*** 홍콩이공대학 조교수 (Hong Kong Polytechnic University · joonoh.seo@polyu.edu.hk)

Received August 7, 2019/ revised August 23, 2019/ accepted August 28, 2019

1. 서론

최근 대형화, 복잡화되는 건설프로젝트를 성공적으로 관리하기 위해서는 대규모 현장에 투입된 수많은 작업자, 중장비, 자재 등의 위치를 파악하는 건설 물체 인식이 매우 중요하다. 실시간으로 수집되는 건설자원들의 위치 데이터는 작업자-장비 간 충돌사고 예방, 작업자의 위험구역 접근 인식 등 안전관리뿐만 아니라 실작업을 분석, 작업/유휴시간 파악, 시간당 작업량 측정 등 생산관리에도 활용될 수 있다(Kim et al., 2019).

이 같은 건설현장 물체 인식의 중요성에 따라, 기존에는 담당자가 건설현장을 직접 방문하고 작업자, 중장비, 자재의 위치를 파악하였다. 하지만 사람이 대규모 현장을 모두 관찰하다 보니 많은 시간과 비용이 요구될 뿐만 아니라 건설 물체(작업자, 중장비, 자재 등)의 위치를 지속적으로 파악하기 어렵다는 한계가 발생하면서, 이를 해결하기 위해 많은 연구자들이 건설현장의 물체를 자동으로 인식하는 기술을 개발하고 있다. 대표적으로, Global Positioning Systems, Radio Frequency Identification, Ultra-Wideband 등의 Internet-of-Things (IoT) 센서를 건설자원에 직접 부착하고 위치 데이터를 취득하는 시스템이 도입되었다. 하지만 IoT 센서를 현장에 투입된 수많은 건설자원에 모두 부착하는 것이 어렵다는 문제가 있어, 최근에는 Closed Circuit Television (CCTV) 카메라를 현장에 설치하고 수집되는 영상 데이터로부터 건설자원의 위치를 파악하는 영상기반 건설현장 물체 인식 시스템을 적용하고 있다. 영상기반 시스템은 IoT 센서를 건설자원에 부착할 필요가 없고 카메라 한 대로 다수의 건설자원을 모니터링할 수 있을 뿐만 아니라, 2016년부터 건설현장 CCTV 설치 및 운용계획 수립이 법제화됨에 따라 영상기반 시스템에 대한 관심이 높아지고 있다(Korea Construction Technology Promotion Act, 2016).

안급한 장점으로 인해 많은 연구자들이 다양한 영상기반 시스템을 개발하였고 실제 현장에서 그 적용 가능성을 보여주었다. 하지만 기존의 연구들은 높은 성능의 영상분석기술을 개발하기 위해서 양질의 학습 데이터베이스(Database, DB)를 구축해야 한다는 한계가 있다. 보다 구체적으로, 인식하고자 하는 건설 물체의 유형과 위치를 학습용 이미지 데이터에 표시하는 Labeling 작업을 일일이 진행해야 하고, 이는 수많은 시간과 노력을 요구한다. 이러한 한계를 보완하기 위해서 본 연구는 학습 데이터 Labeling 작업을 최소화할 수 있는 액티브 러닝(Active Learning)을 활용한 영상기반 건설현장 물체 자동 인식 프레임워크를 제안함을 목표로 한다.

2. 문헌고찰

2.1 액티브 러닝(Active Learning)

액티브 러닝은 대표적인 반지도학습(Semi-supervised Learning)

기법 중 하나이며, 수많은 학습 데이터 중 가장 유의미한 데이터를 우선적으로 선별하고 학습하여 분석 모델의 성능을 점진적으로 최대화하는 과정이다(Chung, 2018). 정보이론(Information Theory)에서 자주 사용되는 엔트로피(Entropy)를 Eq. (1)과 같이 계산하여 학습 데이터에 대한 모델 예측값의 불확실성을 평가하고, 그 불확실성이 가장 높은 데이터를 선택한다(Settles, 2010). 이를 통해 기존의 ‘얼마나 많은 수의 데이터를 학습할 것인가’라는 양적 중심이 아닌 ‘무엇을 먼저 배울 것인가’라는 질적 중심의 학습 DB 및 모델 구축이 가능하다. 액티브 러닝은 학습 데이터의 양을 줄임으로써 DB 구축에 필요한 시간과 노력을 최소화하고 모델의 성능을 최대화할 수 있다.

$$entropy(x) = - \sum_i P(y_i|x) \log P(y_i|x) \quad (1)$$

2.2 영상기반 시스템

건설현장 물체 자동 인식을 위해서 많은 연구들이 진행되었다. 예를 들어, Chi et al.(2009)는 공간 모델링 및 이미지 매칭 기법을 이용하여 건설현장 물체를 인식하는 방법을 제안하였고, Park and Brilakis(2012)는 건설 작업자를 탐지하기 위해서 이미지 배경 추출 방법을 이용하였다. Chi and Caldas(2011)의 경우, 머신러닝 기반의 분류기를 활용하여 건설현장 이미지로부터 작업자, 로더, 백호의 유형을 자동으로 분류하고 위치를 파악하였다. Azar and McCabe(2012a; 2012b)도 Histogram-Of-Gradients 특징을 Support Vector Machine (SVM) 분류기로 학습하고 굴삭기와 덤프트럭을 인식하는 기법을 개발하였다. Memarzadeh et al.(2013)의 연구에서도 SVM을 학습하여 건설 작업자, 굴삭기, 덤프트럭을 탐지하는 방법을 제안한 바 있다. Zhu et al.(2016)은 Particle Filtering 알고리즘을 이용하여 영상에 나타난 건설자원을 추적하였고, Park and Brilakis(2016)과 Kim and Chi(2017) 연구는 인식 및 추적의 상호보완을 통해 영상에서 활발히 움직이는 건설 작업자를 지속적으로 추적하는 방안을 제시하였다.

이 같은 건설자원의 인식 및 추적 결과는 작업 생산성 및 안전성 분석에도 활용되었다. Zou and Kim(2007)은 영상에 나타난 굴삭기의 유휴시간을 분석하는 기술을 제시하였고, Gong and Caldas (2010)은 콘크리트 버킷을 인식함으로써 콘크리트 타설 작업의 사이클시간을 자동 분석하였다. Yang et al.(2014)의 연구에서도 현장 레이아웃 정보를 이용하여 타워크레인의 콘크리트 타설 작업 생산성을 분석하는 방안이 제시되었다. Golparvar-Fard et al.(2013)과 Soltani et al.(2017)은 시공간적 특징(Spatio-temporal Features)을 분석하여 굴삭기의 개별적인 행동(굴착, 회전, 이동 등)을 자동 인식하였다. 이에 더하여, Azar et al.(2013)

과 Kim et al.(2018a)는 굴삭기덤프트럭 간 위치, 거리 등의 상호작용을 분석함으로써 상차작업을 탐지하고 생산성을 분석하는 방법을 제안하였다. 안전성 분석 관련해서는, Chi and Caldas(2012)가 건설장비의 위험구역 접근, 장비 간 충돌 가능성 등 영상기반의 안전 평가 시스템을 개발한 바 있다. Kim et al.(2016)은 분석된 위험정보를 작업자가 빠르고 쉽게 이해할 수 있도록 Fuzzy Inference 모델을 영상분석기술과 연계하였다. Red-Green-Blue Depth 카메라를 활용하여 작업자의 물리적 운동을 분석하고, 이를 토대로 인체공학적 불안정한 행동을 탐지하는 시도도 진행되었다 (Han et al., 2013; Han et al., 2014). Park et al.(2015)는 건설 작업자의 개인보호구 미착용을 자동으로 인식하는 영상기반 프레임워크도 개발하였다.

최근에는 딥러닝 분야의 급격한 발전에 따라 인공지능망(Artificial Neural Network) 모델을 이용하는 영상분석기술들이 개발되고 있다. Fang et al.(2018)과 Kim et al.(2018b)는 건설장비를 인식하기 위해서 지역 기반의 Convolutional Neural Network를 적용하였고, Son et al.(2019)의 연구에서는 Residual Network를 활용하여 다양한 자세의 작업자를 탐지하였다. 이 같은 인식 결과를 토대로 Luo et al.(2018)은 건설자원의 유형, 위치를 추가 분석함으로써 건설작업의 유형을 파악하는 기법을 개발하였고, Kim and Chi(2019)는 Long Short Term Memory (LSTM) 모델을 이용하여 시각적 특성과 행동사이클의 순차적 패턴을 학습함으로써 건설장비의 행동을 자동 인식하는 방법론을 개발하였다. Cai et al.(2019)도 LSTM 모델을 통해 건설장비와 작업자의 위치, 방향, 자세 등을 학습함으로써 다수의 건설자원 간 상호작용을 이해하고 작업유형을 판단하는 프레임워크를 제안하였다.

이처럼 많은 연구들이 진행되었지만 기존의 방법들은 데이터 Labeling 및 학습 DB 구축을 위해 굉장히 많은 시간과 비용을 요구하고 이로 인해 건설현장 물체를 자동으로 인식하는 데 한계가 있다. 따라서 본 연구는 학습 데이터 Labeling 작업을 최소화할 수 있는 액티브 러닝을 활용한 영상기반 건설현장 물체 자동 인식

프레임워크를 제안함을 목표로 한다.

3. 액티브 러닝을 활용한 건설현장 물체 인식 프레임워크

본 프레임워크는 총 3단계로 (1) Unlabeled 데이터의 불확실성 평가, (2) 학습 데이터 선정 및 Labeling, (3) 딥러닝 기반 물체 인식 모델 학습 및 평가로 구성되며(Fig. 1), 단계별 세부내용은 다음과 같다.

3.1 Unlabeled 데이터의 불확실성 평가

본 단계는 액티브 러닝의 핵심과정으로 Unlabeled 학습 데이터에 대한 모델 예측값의 불확실성을 평가하는 것이 목적이다. 이를 위해 먼저 랜덤 샘플링(Random Sampling)을 통해 전체 Unlabeled 이미지 데이터 중 일부 데이터를 추출한다. 추출된 이미지 데이터에 학습된 물체 인식 모델을 적용하고 예측결과의 신뢰도를 계산한 뒤, Eq. (1)을 이용하여 각 이미지에 대한 엔트로피의 총합을 산출한다. 여기서 엔트로피가 높다는 의미는 딥러닝 모델이 이미지에 나타난 물체의 유형과 위치를 파악하기 어렵다는 것을 의미하고, 반대로 엔트로피가 낮다는 의미는 모델의 결과를 신뢰할 수 있음을 나타낸다.

3.2 학습 데이터 선정 및 Labeling

앞 단계의 결과를 토대로 모델의 불확실성(엔트로피의 총합)이 가장 높은 이미지 데이터 N개를 선정하고, 사용자는 Labeling 소프트웨어를 사용하여 선정된 학습용 이미지 데이터의 Labeling 작업을 수행한다(Fig. 2). 이미지 1장당 Labeling 작업을 수행하는 데 평균 약 2분의 시간이 요구되며, 건설분야의 Benchmark dataset (약 430장)의 경우 총 860분 가량이 소요되었다. 이 과정을 통해 모델이 가장 예측하기 어려운 데이터를 선별하고 우선 학습함으로써 Labeling 작업을 최소화하고 모델의 성능을 빠르게 향상할 수 있다.

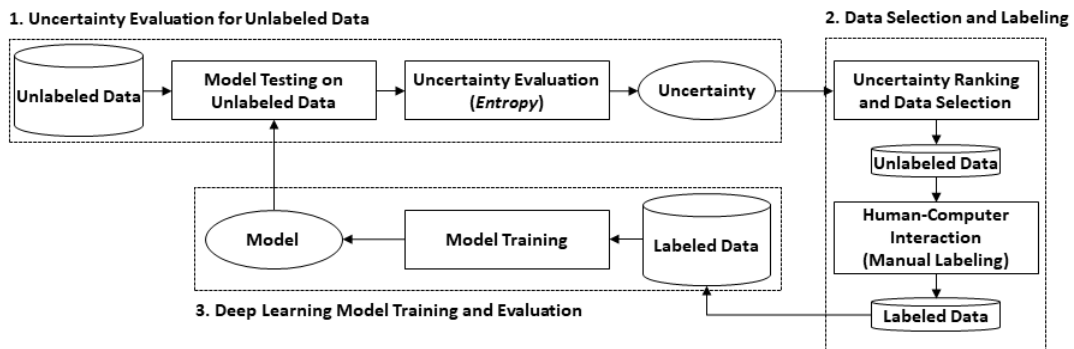


Fig. 1. Active Learning Framework for Construction Object Detection

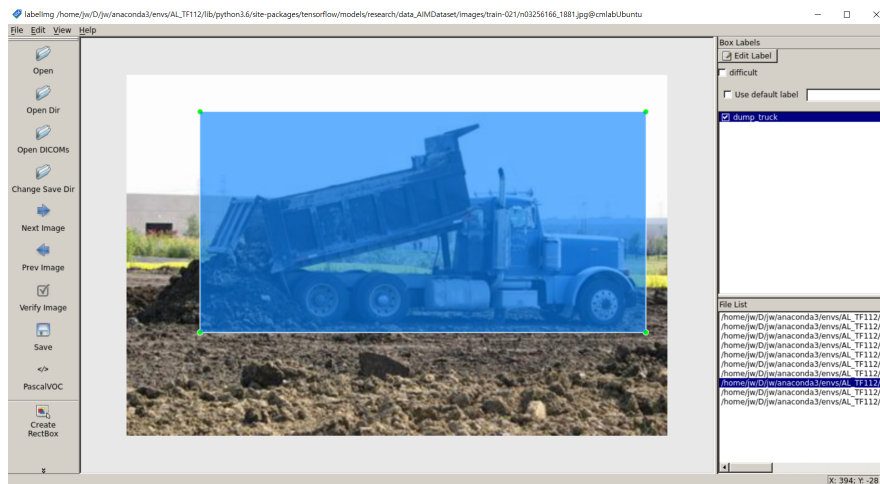


Fig. 2. Example of Image Data Labeling using Labeling Software



Fig. 3. Examples of Construction Object Detection

3.3 딥러닝 기반 물체 인식 모델 학습 및 평가

본 단계의 목적은 Labeling된 이미지 데이터를 활용하여 딥러닝 기반 물체 인식 모델을 학습하고 그 성능을 평가하는 것이다. 특히, 본 연구에서는 Ren et al.(2017)에서 설계한 Faster Region-proposal Convolutional Neural Network (Faster R-CNN) 딥러닝 모델을 건설현장 물체 인식에 맞춤 개발하였다. 해당 모델을 학습하기 위해서 Stochastic Gradient Descent 최적화 알고리즘을 이용하였고 1회당 100번의 학습을 진행하였다. 최종적으로 학습된 모델의 성능을 평가하였고 그 성능이 수렴할 때까지 3-1, 3-2, 3-3단계를 반복 진행하였다.

4. 실험 결과 및 분석

제안한 프레임워크를 검증하기 위해서 Kim et al.(2018b)가 구축한 건설분야의 Benchmark 데이터셋을 이용하여 실험을 수행하였다. 해당 데이터셋은 5개의 건설장비(굴삭기, 덤프트럭, 믹서트럭, 로더, 롤러)에 관한 것이며, 본 연구에서는 총 430개의 이미지 데이터를 활용하였다. 특히, 본 연구에서는 기존의 경험 법칙 (Rule-of-thumb)에 따라 300개(70%)를 Unlabeled 학습 데이터, 130개(30%)를 평가 데이터로 활용하였다. 또한, 초기 영상분석모델을 형성하기 위해서 Google TensorFlow에서 제공하는 기학습된 Faster R-CNN을 활용하였다.

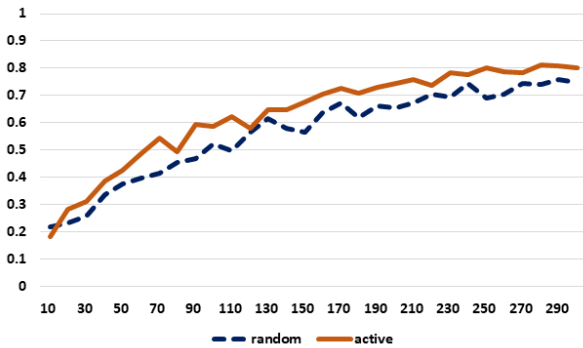


Fig. 4. Performance of Object Detection Model according to the Number of Training Image Data

Fig. 3은 액티브 러닝을 활용한 건설물체 인식 결과 예시를 보여준다. 그림에서 볼 수 있듯이 본 연구에서 제안한 방법은 굴삭기, 덤프트럭, 믹서트럭, 로더, 롤러 등 다양한 건설물체를 성공적으로 인식할 수 있었다. 건설장비별 성능을 살펴보면 믹서트럭 90.1%, 로더 84.5%, 덤프트럭 77.7%, 롤러 74.5%, 굴삭기 73.2% 순으로 나타났다. 보다 구체적으로, 본 연구모델은 덤프트럭의 일부만 이미지 데이터에 나타난 경우에도 인식을 효과적으로 진행하였고 카메라로부터 멀리 존재하여 그 크기가 작은 믹서트럭도 탐지할 수 있었다. 뿐만 아니라 영상분석의 본질적인 한계점이었던 서로 다른 물체가 다양한 각도에서 촬영되어 그 시각적 특성(형태 등)이 변화하더라도 잘 인식하는 것을 볼 수 있다.

또한, 액티브 러닝의 영향을 살펴보기 위해 기존의 방식대로 영상분석모델을 추가 구축하였고 그 실험 결과를 비교 분석하였다. Fig. 4는 학습 데이터 수에 따른 물체 인식 모델의 성능 변화를 나타낸다. 학습 데이터 수가 동일한 조건에서는 기존 방식(Random Learning)보다 본 연구에서 제안한 학습방법(Active Learning)을 사용할 경우 모델의 성능이 높은 것을 알 수 있다. 또한, 모델의 최종 정확도를 살펴보면 기존 방식으로는 약 74%에 머물렀던 반면 액티브 러닝을 진행할 경우 81%까지 성능을 향상할 수 있었다.

5. 결론

본 연구는 학습 DB 구축에 불필요하게 낭비되는 시간과 노력을 최소화하기 위해서 액티브 러닝을 활용한 영상기반 건설현장 물체 자동 인식 프레임워크를 제안하였다. 개발 프레임워크의 성능 평가 및 검증의 목적으로 건설분야 Benchmark 데이터셋을 이용하여 실제 실험을 진행하였다. 실험 결과, 액티브 러닝을 통해 학습한 모델은 다양한 특성을 지닌 건설물체를 성공적으로 인식할 수 있었다. 특히, 기존의 학습 DB 구축 방식으로는 모델의 최대 성능이 74%였던 반면 액티브 러닝을 진행할 경우 81%까지 성능을

향상할 수 있었다. 이는 본 연구결과를 통해 훨씬 더 적은 데이터 수와 반복학습 횟수로도 높은 성능을 가지는 영상분석모델을 개발할 수 있음을 의미하며, 결과적으로 기존에 학습 DB 구축에 요구되는 Labeling 작업을 줄일 뿐만 아니라 총 시간과 비용을 절감할 수 있었다. 향후 연구로는, 모델 성능의 수렴속도 또는 학습시간을 기존 방법과 비교하고 액티브 러닝을 건설현장 물체 인식뿐만 아니라 정지, 이동 등 건설자원의 개별적인 행동과 작업을 파악하는 문제에 적용해볼 수 있다. 또한, Hyperparameter 최적화 등을 통해 알고리즘의 성능을 향상할 수 있다.

감사의 글

본 연구는 국토교통부 국토교통기술촉진연구사업의 연구비지원(19CTAP-C151784-01)에 의해 수행되었습니다.

References

- Azar, E. R. and McCabe, B. (2012a). "Automated visual recognition of dump trucks in construction videos." *Journal of Computing in Civil Engineering*, Vol. 26, No. 6, pp. 769-781.
- Azar, E. R. and McCabe, B. (2012b). "Part based model and spatial-temporal reasoning to recognize hydraulic excavators in construction images and videos." *Automation in Construction*, Vol. 24, pp. 194-202.
- Azar, E. R., Dickinson, S. and McCabe, B. (2013). "Server-Customer interaction tracker: computer vision-based system to estimate dirt-loading cycles." *Journal of Construction Engineering and Management*, Vol. 139, No. 7, pp. 785-794.
- Cai, J., Zhang, Y. and Cai, H. (2019). "Two-step long short-term memory method for identifying construction activities through positional and attentional cues." *Automation in Construction*, Vol. 106, p. 102886.
- Chi, S. and Caldas, C. H. (2011). "Automated object identification using optical video cameras on construction sites." *Computer-Aided Civil and Infrastructure Engineering*, Vol. 26, No. 5, pp. 368-380.
- Chi, S. and Caldas, C. H. (2012). "Image-based safety assessment: automated spatial safety risk identification of earthmoving and surface mining activities." *Journal of Construction Engineering and Management*, Vol. 138, No. 3, pp. 341-351.
- Chi, S., Caldas, C. H. and Kim, D. Y. (2009). "A methodology for object identification and tracking in construction based on spatial modeling and image matching techniques." *Computer-Aided Civil and Infrastructure Engineering*, Vol. 24, No. 3, pp. 199-211.
- Chung, S. (2018). *Bridge damage factor recognition from inspection reports using active recurrent neural network*, Master Thesis, Seoul National University.
- Fang, W., Ding, L., Zhong, B., Love, P. E. D. and Luo, H. (2018). "Automated detection of workers and heavy equipment on cons-

- truction sites: A convolutional neural network approach.” *Advanced Engineering Informatics*, Vol. 37, pp. 139-149.
- Golparvar-Fard, M., Heydarian, A. and Niebles, J. C. (2013). “Vision-based action recognition of earthmoving equipment using spatio-temporal features and support vector machine classifiers.” *Advanced Engineering Informatics*, Vol. 27, No. 4, pp. 652-663.
- Gong, J. and Caldas, C. H. (2010). “Computer vision-based video interpretation model for automated productivity analysis of construction operations.” *Journal of Computing in Civil Engineering*, Vol. 24, No. 3, pp. 252-263.
- Han, S. U., Achar, M., Lee, S. H. and Peña-Mora, F. (2013). “Empirical assessment of a RGB-D sensor on motion capture and action recognition for construction worker monitoring.” *Visualization in Engineering*, Vol. 1, No. 1, pp. 1-13.
- Han, S. U., Lee, S. H. and Peña-mora, F. (2014). “Vision-based detection of unsafe actions of a construction worker: case study of ladder climbing.” *Journal of Computing in Civil Engineering*, Vol. 27, No. 6, pp. 635-644.
- Kim, H. J., Kim, H. K., Hong, Y. W. and Byun, H. R. (2018b). “Detecting construction equipment using a region-based fully convolutional network and transfer learning.” *Journal of Computing in Civil Engineering*, Vol. 32, No. 2, p. 04017082.
- Kim, H. J., Kim, K. N. and Kim, H. K. (2016). “Vision-based object-centric safety assessment using fuzzy inference: monitoring struck-by accidents with moving objects.” *Journal of Computing in Civil Engineering*, Vol. 30, No. 4, p. 04015075.
- Kim, J. W. and Chi, S. H. (2017). “Adaptive detector and tracker on construction sites using functional integration and online learning.” *Journal of Computing in Civil Engineering*, Vol. 31, No. 5, p. 04017026.
- Kim, J. W. and Chi, S. H. (2019). “Action recognition of earthmoving excavators based on sequential pattern analysis of visual features and operation cycles.” *Automation in Construction*, Vol. 104, pp. 255-264.
- Kim, J. W., Chi, S. H. and Seo, J. W. (2018a). “Interaction analysis for vision-based activity identification of earthmoving excavators and dump trucks.” *Automation in Construction*, Vol. 87, pp. 297-308.
- Kim, J., Ham, Y. J., Chung, Y. H. and Chi, S. H. (2019). “Systematic camera placement framework for operation-level visual monitoring on construction jobsites.” *Journal of Construction Engineering and Management*, Vol. 145, No. 4, p. 04019019.
- Korea Construction Technology Promotion Act (2016). *Enforcement decree article 98 and 99, statutes of the Republic of Korea*.
- Luo, X., Li, H., Cao, D., Dai, F., Seo, J. and Lee, S. (2018). “Recognizing diverse construction activities in site images via relevance networks of construction-related objects detected by convolutional neural networks.” *Journal of Computing in Civil Engineering*, Vol. 32, No. 3, p. 04018012.
- Memarzadeh, M., Golparvar-Fard, M. and Niebles, J. C. (2013). “Automated 2D detection of construction equipment and workers from site video streams using histograms of oriented gradients and colors.” *Automation in Construction*, Vol. 32, pp. 24-37.
- Park, M. W. and Brilakis, I. (2012). “Construction worker detection in video frames for initializing vision trackers.” *Automation in Construction*, Vol. 28, pp. 15-25.
- Park, M. W. and Brilakis, I. (2016). “Continuous localization of construction workers via integration of detection and tracking.” *Automation in Construction*, Vol. 72, pp. 129-142.
- Park, M. W., Elsafty, N. and Zhu, Z. (2015). “Hardhat-wearing detection for enhancing on-site safety of construction workers.” *Journal of Construction Engineering and Management*, Vol. 141, No. 9, p. 04015024.
- Ren, S., He, K., Girshick, R. and Sun, J. (2017). “Faster R-CNN: towards real-time object detection with region proposal networks.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 6, pp. 1137-1149.
- Settles, B. (2010). *Active learning literature survey*. Computer Science Technical Report 1648, University of Wisconsin-Madison.
- Soltani, M. M., Zhu, Z. and Hammad, A. (2017). “Skeleton estimation of excavator by detecting its parts.” *Automation in Construction*, Vol. 82, pp. 1-15.
- Son, H. J., Choi, H. C., Seong, H. W. and Kim, C. W. (2019). “Detection of construction workers under varying poses and changing background in image sequences via very deep residual networks.” *Automation in Construction*, Vol. 99, pp. 27-38.
- Yang, J., Vela, P., Teizer, J. and Shi, Z. (2014). “Vision-based tower crane tracking for understanding construction activity.” *Journal of Computing in Civil Engineering*, Vol. 28, No. 1, pp. 103-112.
- Zhu, Z., Ren, X. and Chen, Z. (2016). “Visual tracking of construction jobsite workforce and equipment with particle filtering.” *Journal of Computing in Civil Engineering*, Vol. 30, No. 6, pp. 1-15.
- Zou, J. and Kim, H. (2007). “Using hue, saturation, and value color space for hydraulic excavator idle time analysis.” *Journal of Computing in Civil Engineering*, Vol. 21, No. 4, pp. 238-246.