

논문 2023-18-37

# 다중 도메인 비전 시스템 기반 제조 환경 안전 모니터링을 위한 동적 3D 작업자 자세 정합 기법

## (Dynamic 3D Worker Pose Registration for Safety Monitoring in Manufacturing Environment based on Multi-domain Vision System)

최지동, 김민영\*, 김병학\*

(Ji Dong Choi, Min Young Kim, Byeong Hak Kim)

Abstract : A single vision system limits the ability to accurately understand the spatial constraints and interactions between robots and dynamic workers caused by gantry robots and collaborative robots during production manufacturing. In this paper, we propose a 3D pose registration method for dynamic workers based on a multi-domain vision system for safety monitoring in manufacturing environments. This method uses OpenPose, a deep learning-based posture estimation model, to estimate the worker's dynamic two-dimensional posture in real-time and reconstruct it into three-dimensional coordinates. The 3D coordinates of the reconstructed multi-domain vision system were aligned using the ICP algorithm and then registered to a single 3D coordinate system. The proposed method showed effective performance in a manufacturing process environment with an average registration error of 0.0664 m and an average frame rate of 14.597 per second.

Keywords : Pose Estimation, Point Cloud, 3D Registration, Deep Learning, RGB-D

### 1. 서론

현대 제조 산업에서는 공장 자동화를 통해 생산성을 향상하고 공정의 효율성을 높이며 인적 노동의 부담을 줄이기 위하여 작업자와 협력이 가능한 갠트리 로봇과 협동 로봇 활용시스템 개발이 경쟁적으로 진행되고 있다. 또한 작업자와 로봇 간의 협력이 가능한 다양한 제조 공정 유형이 새롭게 도입됨에 따라 협력 제조 공정에서 작업자와 로봇은 작업 공간을 공유한다. 작업 공간의 공유는 충돌과 같은 안전 문제를 발생시키며 이를 방지하고자 첨단 로봇들은 안전 프로토콜 및 초음파 및 적외선 센서를 이용한 안전 대응 기술을 갖추고 있지만, 협력 제조 공정과 환경의 다양성에 따라

서 예기치 못한 문제가 발생 된다.

제조 공정 환경에서의 안전 문제는 매우 중요하며 이를 해결하기 위해 다중 도메인 비전 시스템의 시각적 데이터와 인공지능 기술을 응용하여 객체 탐지 (Object Detection), 객체 추적 (Object Tracking), 자세 추정 (Pose Estimation) 등과 같은 다양한 기법들이 활발하게 연구되고 있다. 자세 추정 기법은 이미지나 비디오 데이터에서 사람이나 다른 객체의 자세 (체형, 관절 위치) 및 동작을 추정하고 분석하는 기법으로써, 일반적인 객체 탐지 및 객체 추적과는 다르다. 자세 추정의 성능은 추정된 관절 위치가 실제 관절 위치와 일치하는 비율에 따라 정확도 (Accuracy)와 사람이나 다른 객체의 추정된 경계 상자 (Bounding Box)와 실제 경계 상자 간의 겹치는 영역 (Intersection over Union) 및 유클리드 거리 (Euclidean distance)를 사용하여 추정된 관절 위치와 실제 관절 위치 간의 평균 거리를 계산하여 평가될 수 있다.

기존에는 자세 추정 기법을 위해 2차원 영상 (RGB) 데이터와 2차원 영상과 깊이 (Depth) 센서를 결합한 RGB-D 데이터를 활용한 연구가 진행되었지만, 최근에는 2차원 영상과, 3차원 영상, 라이다 (LiDAR)를 결합한 다중 도메인 비전 시스템을 활용한 자세 추정 기법 연구가 활발하게 진행되고 있다.

RGB 카메라는 색상 정보를 통해 객체 인식에 유리하나, 깊이 데이터의 부재로 인해 공간적 인식에는 한계가 있다. RGB-D 카메라는 깊이 정보를 제공하며 3차원 환경 인식 성능도 정확한 3차원 포인트 클라우드 데이터를 얻을 수 있지만, 색상 정보 데이터가 부족하고, 비용이 상당히 높다는 단점이 있다. 이러한 센서들의 단점은 외부 보정 기법을 통

\*Corresponding Author (minyoungkim@knu.ac.kr, bhkim81@kitech.re.kr)  
Received: Sep. 29, 2023, Revised: Oct. 19, 2023, Accepted: Nov. 20, 2023.  
J. D. Choi: Korea Institute of Industrial Technology, Kyungpook National University (Research Student)  
M. Y. Kim: Kyungpook National University (Prof.)  
B. H. Kim: Korea Institute of Industrial Technology (Senior Researcher)  
\* This study has been conducted with the support of the Korea Institute of Industrial Technology (KITECH). This work was supported by the Industrial Technology Innovation Program (20016970, Development of Cooperative Robot SI (System Integration) Service based on Safety Intelligence) funded by the Ministry of Trade Industry & Energy (MOTIE, Korea), This work was supported by the Industrial Technology Innovation Program(20018288, Development of 150-ton crawler crane with vision-based intelligent safety management system) funded by the Ministry of Trade Industry & Energy(MOTIE, Korea), and the Core Research Institute Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2021R1A6A1A03043144) and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2022R1A2C2008133).

해 상호 보완될 수 있으며 각 센서의 장점을 최대한 활용할 수 있다. 실시간 동적 작업자의 자세 정합을 수행하는 안전 대응 응용시스템은 다중 도메인 비전 시스템 간의 외부 보정이 매우 중요하다. 외부 보정을 통해 다양한 센서의 데이터를 융합하고 정확도를 향상시킬 수 있다. 다중 도메인 비전 시스템은 2차원 영상, 3차원 영상 및 3차원 포인트 클라우드 데이터를 모두 가지고 있으므로 기존의 2차원 영상 데이터만을 사용하는 외부 보정 기법과 달리 3차원 포인트 클라우드 데이터를 이용하여 다중 도메인 비전 시스템을 보정할 수 있다. 정확한 다중 도메인 비전 시스템의 외부 매개변수를 추정하기 위해 다양한 연구가 수행되었다. 기존의 외부 보정 기법은 2차원 영상 기반 보정 보드를 이용하여 특징점을 추출하는 기법이 있다. 보정 보드를 이용하는 경우 보정 보드를 들고 이동하면서 2차원 영상 데이터를 촬영해야 하고, 보정 보드가 변경되면 보정 기법도 변경해야 한다는 단점이 있다. 3차원 포인트 클라우드 기반에는 지정되지 않은 객체의 모양을 활용하는 기법이 있다. 3차원 포인트 클라우드 기반 외부 보정 기법은 객체의 모양을 촬영한 후 특징점을 찾고 특징점을 이용하여 외부 보정을 수행하므로 보정 보드가 필요하지 않다는 장점이 있다.

본 연구는 특별한 보정 보드 없이 제조 공정 환경에서 갠트리 로봇을 특징점으로 사용하여 다중 도메인 비전 시스템으로 촬영한 3차원 포인트 클라우드 데이터를 통해 외부 보정을 수행하고 추정된 외부 매개변수로 동적 3D 작업자의 자세 정합 기법을 제안하였다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구를 정리한다. 3장에서는 디지털트윈기반 제조 환경 안전 모니터링을 설명하고 4장에서는 제안하는 알고리즘을 설명한다. 5장에서는 실험 방법 및 결과를 보여주고 6장에서 논문을 마무리한다.

## II. 관련 연구

### 1. 2차원 영상 기반 자세 추정 기법

2차원 영상 기반의 자세 추정은 사람이나 다른 객체의 체형과 동작을 추정하고 분석하는 기법으로써 기존에는 기계 학습 방법과 hand-crafted 특징들을 중심으로 연구가 진행되었다. 기계 학습 및 hand-crafted 특징 기반의 DPM (Deformable Part Models) 알고리즘은 객체를 여러 부위로 나누고 부위 간의 관계를 모델링 한다. 각 부위는 독립적으로 인식하고 상대적인 위치와 모양 정보를 활용하여 자세를 추정한다 [1, 2]. 하지만 기계 학습 방법 및 hand-crafted 방법들은 복잡한 자세와 다양한 배경, 조명 등의 변화에서 성능의 한계가 있었다. 이러한 한계점을 해결하기 위해 최근에는 DeepPose, Stacked Hourglass Network, OpenPose 와 같은 CNN (Convolution Neural Network) 기반의 사전학습 모델과 기계 학습 및 hand-crafted 특징 기반 모델들이 개발되어 성능을 크게 향상 시켰다 [3-5]. CNN 기반의 자세 추정 모델은 이미지에서 객체의 관절 위치를 직접 예측한다.

DeepPose 자세 추정 모델은 CNN을 이용하여 전체 이미지에서 각 관절의 위치를 회귀적으로 예측하며 자세 추정 문제를 회귀 문제로 해결했다. Stacked Hourglass Network 자세 추정 모델은 hourglass 모듈을 융합하여 다양한 스케일의 특징을 학습하고 세밀한 위치 정보와 다양한 자세 변화를 추정할 수 있었다. OpenPose 자세 추정 모델은 실시간으로 다중 객체의 자세를 추정하기 위해 제안되었다. PCM (Part Confidence Maps), PAFs (Part Affinity Fields) 두 가지 기법을 제안하며 2차원 영상 데이터에서 bottom-up 방식을 이용하여 각 관절 위치의 확률을 예측하며 사람의 다른 관절 사이의 연관성을 추정한다. PCM는 각 픽셀이 특정 관절에 속할 확률을 나타내는 2D 확률 맵을 생성하며 PAFs는 각 관절 쌍의 위치와 방향을 고려하여 각 관절들이 한 객체에 속하는지 판단한다. 두 가지 기법으로 OpenPose는 실시간 환경에서도 다중 객체의 관절 위치를 정확하게 추정할 수 있다. HRNet (High-Resolution Network)는 고해상도의 특징 맵을 유지하면서 서로 다른 스케일의 특징을 병렬로 처리하는 자세 추정 모델이다. 대다수의 CNN 기반 모델들은 네트워크의 깊이가 증가함에 따라 특징 맵의 해상도가 감소하지만, HRNet은 모든 단계에서 고해상도를 유지한다. 다양한 스케일의 특징을 병렬로 추출한 뒤 이를 다시 고해상도 특징 맵으로 통합한다. HRNet은 세밀한 고 레벨의 특징 정보와 넓은 범위의 저 레벨의 맥락 정보를 동시에 통합할 수 있어 높은 정확도 성능을 보여주고 있다 [6]. 최근 이러한 모델에서 과생되어 Self-Attention 기법을 적용한 새로운 모델에 대한 연구도 활발히 이루어지고 있다 [7-10].

### 2. 3차원 영상 기반 자세 추정 기법

3차원 영상 데이터를 사용하여 객체의 자세를 정확하게 추정하려는 연구들이 진행되고 있다. Shotton 등의 연구에서는 Decision Forest 알고리즘을 사용하며 다수의 Decision Trees로 구성된 앙상블 (Ensemble) 모델로 각 트리는 3차원 영상의 각 픽셀에 대응하는 객체의 관절 위치를 확률을 예측했다 [11]. Zimmermann 등의 연구에서는 3차원 사람 포즈를 추정하기 위해 2차원 영상 이미지와 3차원 영상 이미지를 사용하는 CNN 모델을 제안했다 [12]. Bashirov 등의 연구에서는 SMPL-X (Skinned Multi-Person Linear-expressive) 모델을 사용하여 사람의 자세 및 동작을 예측하고 3차원 랜드마크로부터 사람 관절 각도를 추정하고 얼굴의 표정과 턱 자세를 추출한다 [13]. Veges 등의 연구에서는 3차원 자세를 추정하기 위해 깊이 예측 네트워크를 사용한다. 2차원 영상 입력 이미지에서 각 관절에 해당하는 좌표를 추정하고, 3차원 영상에 2차원 관절 좌표를 회귀 (Regression) 하여 사람의 3차원 관절 좌표를 추정한다 [14].

### 3. 3차원 포인트 클라우드 기반 자세 추정 기법

3차원 포인트 클라우드 데이터를 활용한 자세 추정 연구들은 현재 객체의 관절 위치와 구조를 더욱 정확하게 추정하기 위해 진행되고 있다. 포인트 클라우드 데이터는 3차원

공간에서의 점들의 집합으로써 물체의 외형과 구조를 형상화 할 수 있다. V2V-PoseNet [15]은 단일 깊이 맵에서 정확한 3차원 손과 사람의 자세를 추정한다. 깊이 맵을 3차원 볼륨으로 변환하고 사람의 자세를 추정한다. V2V-PoseNet은 3D convolution layers로 구성되어 있으며, 각 복셀 (voxel)이 사람의 관절 위치에 맵핑될 수 있게 한다. 단일 깊이 이미지만을 입력으로 사용하여 사람의 여러 관절의 3D 좌표를 예측하며 다양한 동작과 환경에서의 높은 정확도를 보였다. DensePose [16]는 다양하고 복잡한 환경에서의 사람 자세를 추정한다. DensePose는 깊이 맵의 사람에 해당하는 픽셀을 3D 모델에 맵핑한다. 이 방법은 fully convolutional network를 사용하여 각 픽셀에 대한 관절 위치와 3차원 좌표를 예측한다. 다양한 동작과 사람의 모양, 크기, 외관에 대한 자세 추정 높은 정확도를 보여준다. LCR-Net [17]은 2차원 영상에서 사람의 3차원 자세를 추정한다. LCR-Net은 사람의 각 관절을 로컬라이제이션 (Localization) 하고, 각 부위의 클래스를 분류 (Classification)하며, 3차원 회귀 모델을 사용하여 정확한 3차원 관절 좌표를 예측한다.

4. 제조 산업 환경 작업자 자세 추정 기법

최근 제조 산업 분야에서는 안전하고 효과적이며 유연한 제조를 위하여 작업자와 협동 로봇 간의 협업 작업에 대해 디지털 트윈 (Digital Twin) 기술의 적용에 대한 연구가 활발하게 진행되고 있다. Maruyama 등의 연구에서는 광학 모션 캡처 방식인 OptiTrack [18]을 활용해 작업자의 옷, 장갑, 신발 등 37개의 마커를 부착하고 3차원 면 (Mesh)으로 모델링하여 작업자의 자세를 추정하는 기법을 제안하였다 [19]. 그러나 이 기법은 37개의 마커의 데이터를 Modbus 통신을 이용하여 처리할 때, 작업 속도가 급격하게 증가하거나 감소하는 경우 자세 추정 정확도가 떨어진다. 또한 가려짐을 방지하기 위해 여러 카메라의 배치가 필요하며 구현 환경에 따라 제약이 발생한다. Weistroffer 등의 연구에서는 DHM (Digital Human Model)을 모델링하기 위해 허리, 다리, 팔 등 최소 5개의 모션 캡처 마커를 부착하여 자세를 추정하고 관절 위치 추정 정확도 성능을 향상하기 위해 손목, 머리 등과 같은 부위는 IMU 센서를 활용하여 관절 위치 추정의 높은 정확도 성능을 보여주고 있다 [20]. 그러나 기존 기법들은 제조 산업 환경에서 작업자의 신체에 광학 모션 캡처 마커를 부착해야하는 번거로움과 작업 속도에 따라 자세 추정 정확도 성능이 급격히 저하되는 문제로 인해 실제 환경에서 기술 적용상의 한계사항이 존재한다.

5. 다중 도메인 비전 시스템 외부 보정 기법

각 센서의 장단점들을 상호 보완하며 장점을 최대한 활용할 수 있도록 연구가 활발하게 진행되고 있다. Choi 등의 연구에서는 아크릴 보정 보드를 이용하여 열화상 카메라와 라이다 간 상대적인 위치와 방향을 추정하는 외부 보정 기법을 제안했다 [21]. Park 등의 연구에서는 다각형 보정 보드를 이용하여 라이다와 RGB 카메라 간 정확한 외부 보정

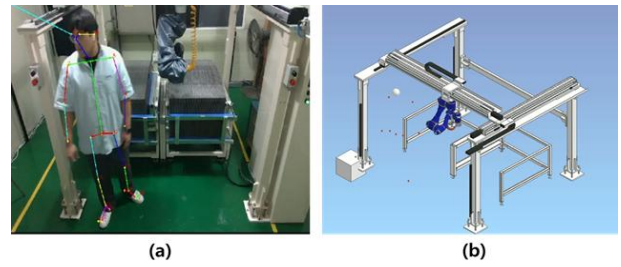


그림 1. 디지털 트윈 결과 (a)동적 작업자 자세 추정 결과 (b) 디지털 트윈 영상

Fig. 1. Digital twin results (a) Dynamic worker posture estimation results (b) Digital twin image

기법을 제안했다 [22]. Liu 등의 연구에서는 다중 RGB-D 카메라의 외부 보정을 위해 기존의 보정 보드가 아니라 Descriptor-Based Pattern을 사용하여 다중 RGB-D 카메라의 각각의 카메라에 대한 특징을 추출하고 카메라 간의 상대적인 위치와 방향을 추정했다 [23]. Lee 등의 연구에서는 다중 RGB-D 카메라의 외부 보정을 위해 보정 보드를 사용하지 않고 각 카메라에서 인지된 사람을 특징 매칭 (feature matching) 알고리즘을 통해 각 RGB-D 카메라 간의 상대적인 위치와 방향을 추정했다 [24].

III. 디지털 트윈기반 제조 환경 안전 모니터링

디지털 트윈은 실제 객체, 시스템을 가상의 모델로 생성하여 실시간으로 데이터를 수집하여 모니터링, 분석, 관리할 수 있다. 그림 1 (a)와 같은 제조 공정을 생성하기 위해 갠트리 로봇, 협동 로봇의 도면을 활용하여 (b)와 같이 3차원 모델링을 생성하였다. 디지털 트윈 기반 제조 환경 안전 모니터링 시스템 연동 설계를 위하여 그림 2 와 같이 ICD (Interface Control Document)를 설계하였다. 제조 공정 내 갠트리 로봇, 협동 로봇 및 동적 작업자의 3차원 관절 위치 정보 시각화 기능을 위해 데이터들은 PLC, (Programmable Logic Controller), Modbus 와 같은 이더넷 (Ethernet) 통신 프로토콜을 활용하여 구성하였다. 디지털 트윈을 통해 공정 안전 관리자는 제조 공정 환경을 실시간으로 모니터링 할 수 있다.

IV. 동적 3D 작업자 자세 정합

1. 다중 도메인 비전 센서 데이터 정합

단일 비전 시스템을 사용할 경우, 공정 내 갠트리 로봇과 협동 로봇에 의해 동적 작업자의 관절 일부가 가려질 수 있다. 가림 현상으로 단일 비전 시스템은 로봇과 동적 작업자 사이의 상호 작용 및 공간적 제약을 정확하게 파악하기 어렵다. 이러한 문제점으로 제조 공정 안전 모니터링에서 성능의 한계 사항을 보여준다. 이러한 한계점을 해결하기 위해 제안한 방법은 그림 3 과 같이 다중 도메인 비전 시스템으로부터 획득한 시점이 다른 3차원 포인트 클라우드 데이터

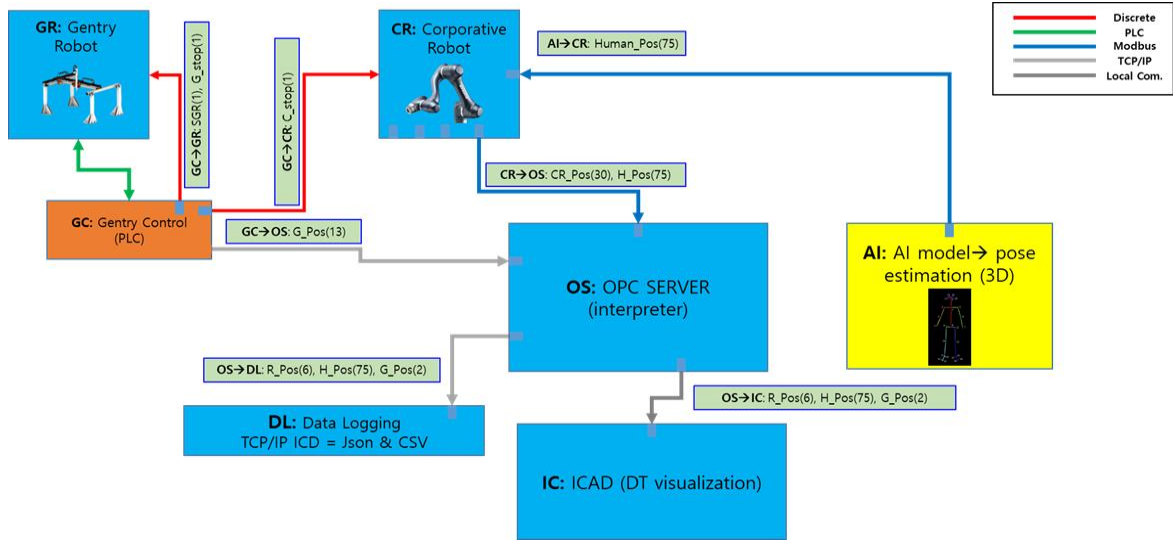


그림 2. 제조 환경 안전 모니터링을 위한 디지털 트윈 ICD 개념도  
 Fig. 2. Digital twin ICD concept for safety monitoring in manufacturing environments

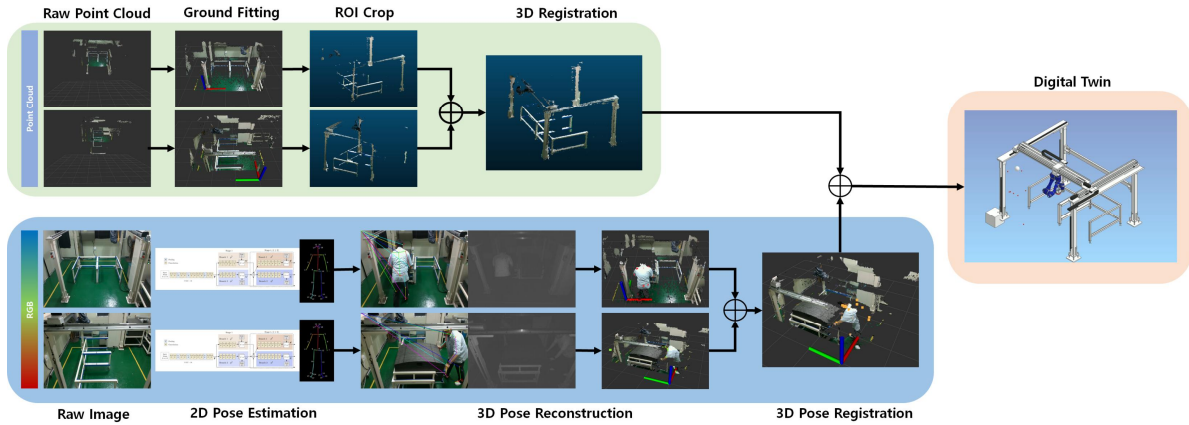


그림 3. 다중 도메인 비전 시스템 기반 제조 환경 안전 모니터링을 위한 동적 3D 작업자 자세 정합 기법 개념도  
 Fig. 3. Conceptual diagram of dynamic 3D worker pose registration for safety monitoring in manufacturing environments based on multi-domain vision system

에서 지면 피팅 (Ground Fitting)을 통해 지면과 3차원 포인트 클라우드 데이터가 정렬된다. 갠트리 로봇에 해당하는 관심 영역 (ROI, Region of Interest)을 지정하고 관심 영역에 해당하는 데이터를 통해 다중 도메인 비전 시스템을 외부 보정한다. 제안한 방법이 기존 다중 도메인 비전 시스템의 외부 보정 방법 [24]과 차별화 된 점은, 카메라 사이의 거리 (Base Line)와 수렴 각도 (Convergence Angle)에 따른 제약이 없고 특별한 보정 보드 없이 제조 공정 내 갠트리 로봇의 3차원 영상만을 이용하고 다중 도메인 비전 시스템 외부 보정 다중화하기 위해 ICP (Iterative Closest Point) 알고리즘 [25]을 활용하여 시점이 다른 3차원 포인트 클라우드 데이터 정합을 위한 회전 이동 행렬을 추정하는 방식을 사용하는 부분에서 기존의 기술과 차별화된다.

$$e = \sum_{i=1}^K ((RC_{Sp} + T - C_{Fp})n_i)^2. \quad (1)$$

수식 (1)에서  $R$ ,  $T$ 는 회전 및 이동 행렬이며  $C_{Sp}$ ,  $C_{Fp}$ 는 다 시점에서 획득된 다 시각 라이다의 3차원 포인트 클라우드를 의미한다. 법선 벡터인  $n_i$ 를 기준으로 3차원 포인트 클라우드 데이터가 정렬된다.  $e$ 는 두 공간 데이터의 각 포인트 간 최소 근접 데이터 간의 오차 합산 값으로  $e$ 값이 최소가 되는 반복 과정을 통하여 최종 정합을 위한  $R$ ,  $T$  행렬을 추정한다.

### 2. 동적 작업자 자세 추정 및 정합

제조 공정 내 동적 작업자의 3차원 위치 및 자세를 추정하기 위해 Open Pose 모델을 사용했다. Open Pose는 PAFs



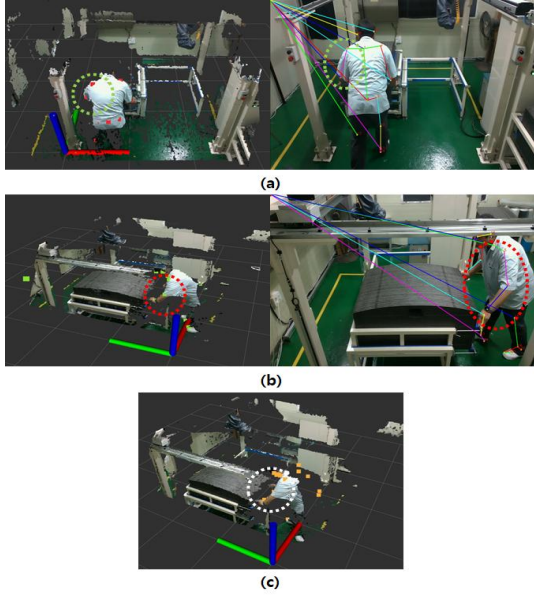


그림 4. 동적 작업자 자세 추정 및 정합 결과  
 (a) 정면 단일 비전 시스템 (b) 옆면 단일 비전 시스템  
 (c) 다중 도메인 비전 시스템

Fig. 4. Title of Figure Dynamic worker pose estimation and registration results  
 (a) frontal single vision system (b) side single vision system (c) multi-domain vision system

와 PCM을 활용하여 다중 객체의 자세를 실시간으로 추정할 수 있는 bottom-up 방식의 자세 추정 모델이다. 그림 3과 같이 2차원 영상에서 동적 작업자의 관절의 위치 정보를 추정하고, 추정된 2차원 관절 위치 정보는 3차원 영상에 맵핑되어 각 픽셀에 대응하는 동적 작업자의 관절 좌표를 3차원 좌표로 재구성 (Reconstruction)한다. 다 시점에서 획득된 동적 작업자의 3차원 관절 정보를 정합하기 위해 수식 (1) 에서 도출된 회전 이동 행렬  $R$ ,  $T$ 를 이용하여 다 시점 3차원 관절 정보를 정렬할 수 있다. 다 시점 3차원 관절 정보를 정렬 하였으나 하나의 3차원 관절 정보로 데이터를 정합하는 과정이 필요하다. 3차원 데이터 정합을 위해 수식 (2) 방식을 이용하여 다 시점 3차원 관절 정보 간의 유클리디안 거리 (Euclidean Distance)  $d$ 를 계산한다.  $x_{l_i}$ ,  $y_{l_i}$ ,  $z_{l_i}$ 와  $x_{r_i}$ ,  $y_{r_i}$ ,  $z_{r_i}$ 는 다 시점에서 획득된 3차원 관절 위치 정보 데이터 ( $x, y, z$ )를 의미한다.  $F_i$ ,  $S_i$ 는 다 시점에서 획득된 동적 작업자의  $i$ 번째 3차원 관절 좌표를 의미한다. 정렬된 다 시점의 3차원 관절 정보 데이터에서 수식 (3)의 방식을 이용하여 유클리디안 거리  $d$ 를 기반으로 임계값인  $threshold$  보다 작으면 3차원 관절 좌표  $P_{regi}$ 로 정합한다.

$$d(F_i, S_i) = \sqrt{(x_{l_i} - x_{r_i})^2 + (y_{l_i} - y_{r_i})^2 + (z_{l_i} - z_{r_i})^2}, \quad (2)$$

$$P_{regi} = \left\{ \frac{1}{n} \sum_{i=1}^n F_i \mid d(F_i, S_i) < threshold \right\}. \quad (3)$$

그림 4 (a) 와 같이 단일 비전 센서 시스템의 한계를 보여주며, 단일 비전 센서 시스템을 사용할 경우 동적 작업자의 위치 변화에 따라 오른쪽 팔의 3차원 관절의 위치를 정확하게 추정하는데 어려움이 있다. (b)와 같이 다른 시점의 단일 비전 센서 시스템에서는 동적 작업자 오른쪽 팔의 3차원 관절 위치를 보다 정확하게 추정할 수 있음을 보여준다. (c)와 같이 제안된 다중 도메인 비전 센서 시스템 기반 동적 3D 작업자 자세 정합 기법을 사용하였을 때, 제조 공정 내 동적 작업자의 3차원 자세 추정 성능이 향상된 효과를 보여준다.

### V. 실험 결과

본 논문에서 제안한 다중 도메인 비전 시스템 기반 제조 환경 안전 모니터링을 위한 동적 3D 작업자 자세 정합 기법의 성능을 평가하고자 그림 5 (a) 와 같이 제조 공정 내 배치된 갠트리 로봇과 협동 로봇을 활용하는 작업 공간 내 작업자의 위치를 추정하기 위해 RGB-D 센서로 구성된 다중 도메인 비전 시스템 간  $90^\circ$  의 시야 (FOV, Field Of View)가 서로 겹치도록 실험 환경을 구성하였다. (b)와 같이 갠트리 로봇과 협동 로봇의 작업 공간 범위에 해당하는 노란색 사각형 범위 내에서 (a)~(c) 위치로 움직이며 동적 작업자의 3차원 관절 위치 정보를 획득한다. 성능 검증을 위해 3개의 평가 항목인 평균 정합 오차 (Registration error), 관절 추정 개수, 평균 수행 시간으로 평가했다. 평균 정합 오차 추정은 정면의 단일 비전 시스템의 동적 작업자의 3차원 관절 위치 정보를 기준으로 정렬된 다른 시점의 단일 비전 시스템의 동적 작업자의 3차원 관절 위치 간에 수식 (4)의 방식으로 다중 도메인 비전 시스템에서 추정된 3차원 동적 작업자 관절 의 평균 정합 에러  $E$  를 계산한다.

$$E = \frac{1}{n} \sum_{i=1}^n d(F_i - S_i). \quad (4)$$

수식 (4)에서  $n$ 은 다른 시점의 단일 비전 시스템의 동작 작업자의 3차원 관절에 해당하는 포인트 클라우드 데이터의 개수이다.  $F_i$ ,  $S_i$ 는 다중 도메인 비전 시스템의 동적 작업

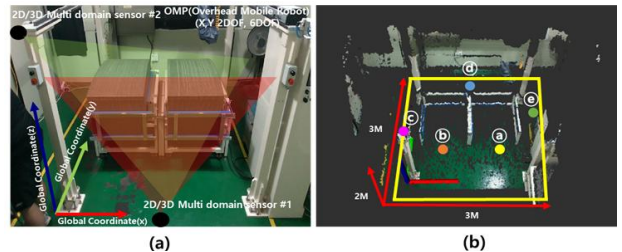


그림 5. 실험 환경 및 실험 데이터 획득 예시  
 (a) 실험 환경 (b) 실험 데이터 획득 방법

Fig. 5. Example of experimental environment and experimental data acquisition  
 (a) Experimental environment (b) Method of acquiring experimental data

표 1. 다 시점 동적 작업자의 3차원 관절 평균 정합 오차 표  
Table 1. 3D joint average registration error of multi-view dynamic worker

Data	Registration Error (m)	Position
1	0.0694	(a)
2	0.0644	(b)
3	0.0580	(c)
4	0.0713	(d)
5	0.0690	(e)
Average error(m)	0.0664	

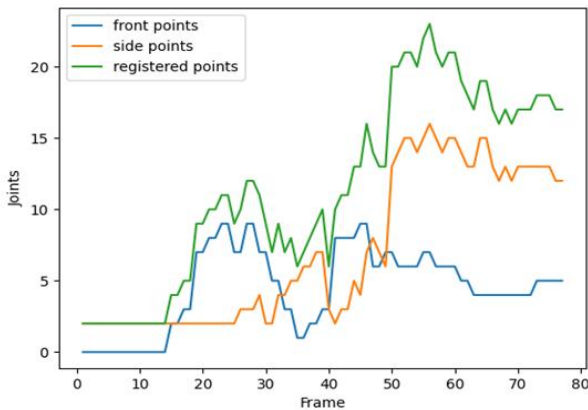


그림 6. 다 시점 동적 작업자 3차원 관절 추정 개수 실험 결과  
Fig. 6. Multi-view dynamic worker 3D joint number estimation experiment results

자의 3차원 관절에 해당하는 포인트 클라우드이며,  $d(F_i - S_i)$ 는 3차원 관절의 유클리디안 거리를 의미한다.

표 1은 다중 도메인 비전 시스템 기반 동적 작업자의 3차원 관절의 평균 정합 오차를 측정된 결과이다. 그림 5 (b)와 같이 작업 공간 범위에 해당하는 노란색 사각형 범위 내에서 동적 작업자가 (a)~(e) 해당 위치로 움직이며, 실험 데이터를 취득했다. 5개의 각 위치에서 추정된 동적 작업자의 3차원 관절 정보 데이터의 평균 정합 오차는 각 위치에서 0.1m 이내이며, 5개 위치 모두에 대한 평균 정합 오차는 0.0664m이다. 그림 6과 같이 제조 공정 내 동일한 위치에서 단일 도메인 비전 시스템과 제안한 다중 도메인 비전 시스템을 이용한 동적 작업자의 3차원 관절 추정 개수에서 단일 도메인 비전 시스템보다 제안한 방법이 동적 작업자의 3차원 관절 추정 개수 성능이 향상된 결과를 보여주었다.

국제 표준화 기구 (ISO)는 협동 로봇의 안전성을 확보하기 위한 표준 규격인 ISO/TS 15066을 통해 로봇과 작업자 간의 상호 작용 시 안전을 보장하기 위한 방안의 하나로 로봇과 작업자의 속도에 따라 보호 이격 거리를 동적으로 계산할 수 있도록 수식 (5)을 제안했다 [26].

$$S_h = 1.6T_s + 1.0T_r \quad (5)$$

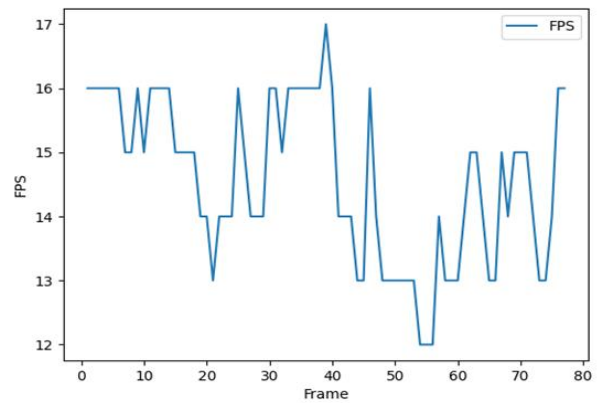


그림 7. 제안한 방법의 평균 인식 시간 결과  
Fig. 7. Average recognition time results of the proposed method

수식 (5)에서  $T_r$ 는 로봇이 위험을 감지하고 멈추는 시간이며,  $T_s$ 는 작업자가 위험을 인지하고 반응하는 시간이다. ISO/TS 15066 표준에 따라 작업자의 평균 이동 속도  $1.6m/s$  협동 로봇의 최대 이동 속도  $1.0m/s$  으로부터 추정된 보호 이격 거리  $S_h$ 는  $0.55m$ 로 산출된다. 그림 7은 제안된 방법의 평균 인식 시간을 나타내며, 평균 인식 시간 14.597 FPS (Frame Per Second)로 보호 이격 거리  $S_h$ 와 평균 인식 시간에 따라 제안된 방법의 거리 변화 인식 해상도는  $0.11m$ 로 갱신된다.  $1.5m$  거리에서는 갠트리 로봇 및 협동 로봇의 작업 속도 조절할 수 있으며,  $1.0m$  이내에서는 비상 정지를 수행하도록 안전율을 산정할 수 있다.

## VI. 결론

본 논문에서는 다중 도메인 비전 시스템 기반 제조 환경 안전 모니터링을 위한 동적 3D 작업자 자세 추정 방법을 제안하였다. 제조 공정 환경에서 동적 작업자의 3차원 자세 추정을 위해서 Open Pose 모델을 활용하여 2차원 자세를 추정하였다. 추정된 2차원 자세를 RGB-D 센서를 이용해 3차원으로 재구성 하였다. 다중 도메인 비전 시스템 동적 작업자의 3차원 관절 위치 정보는 제안한 방법으로 새로운 3차원 관절로 정합 한다. 제안한 방법은 평균 정합 오차가  $0.0664m$ , 평균 수행 시간이 14.597 FPS로 실시간으로 효율적으로 동작할 수 있음을 확인하였다.

향후 계획으로는 다중 도메인 비전 시스템을 통해 추정된 동적 작업자의 3차원 관절 위치 정보를 기반으로 갠트리 로봇 및 협동 로봇 간 상대적인 거리를 추정하여 제조 공정 내 로봇과 동적 작업자 간 발생할 수 있는 위험한 상황을 판단 및 제어 방법에 관한 연구를 수행할 계획이다.

## References

- [1] P. F. Felzenszwalb, D. P. Huttenlocher, "Pictorial Structures for Object Recognition," *International Journal of Computer Vision*, Vol. 61, No. 1, pp. 55 - 79, 2005.
- [2] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, D. Ramanan, "Object Detection with Discriminatively Trained Part-based Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 9, pp. 1627 - 1645, 2010.
- [3] A. Toshev, C. Szegedy, "DeepPose: Human Pose Estimation Via Deep Neural Networks," *IEEE Conference on Computer Vision and Pattern Recognition*, pp.1653 - 1660, 2014.
- [4] A. Newell, K. Yang, J. Deng, "Stacked Hourglass Networks for Human Pose Estimation," *European Conference on Computer Vision*, pp. 483-499, 2016.
- [5] Z. Cao, G. Hidalgo, T. Simon, S. E. Wei, Y. Sheikh, "Openpose: Realtime Multi-person 2d Pose Estimation Using Part Affinity Fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 43, No. 1, pp. 172 - 186, 2019.
- [6] K. Sun, B. Xiao, D. Liu, J. Wang, "Deep High-resolution Representation Learning for Human Pose Estimation," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5693 - 5703, 2019.
- [7] Y. Zhou, X. Wang, X. Xu, L. Zhao, J. Song, "X-hrnet: Towards Lightweight Human Pose Estimation with Spatially Unidimensional Self-attention," *IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1 - 6, 2022.
- [8] W. Zhao, W. Wang, Y. Tian, "Graformer: Graph-oriented Transformer for 3d Pose Estimation," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20438 - 20447, June 2022.
- [9] Q. Zhao, C. Zheng, M. Liu, P. Wang, C. Chen, "PoseFormerV2: Exploring Frequency Domain for Efficient and Robust 3D Human Pose Estimation," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8877 - 8886, 2023.
- [10] J. Cai, H. Liu, R. Ding, W. Li, J. Wu, M. Ban, "Htnet: Human Topology Aware Network for 3d Human Pose Estimation," *arXiv preprint arXiv:2302.09790*, 2023.
- [11] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. "Real-time Human Pose Recognition in Parts from Single Depth Images," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1297 - 1304, 2011.
- [12] C. Zimmermann, T. Welschhold, C. Dornhege, W. Burgard, T. Brox, "3D Human Pose Estimation in RGBD Images for Robotic task Learning," *IEEE International Conference on Robotics and Automation*, 2018.
- [13] R. Bashirov, A. Ianina, K. Isakov, Y. Kononenko, V. Strizhkova, V. Lempitsky, A. Vakhitov, "Real-Time RGBD-Based Extended Body Pose Estimation," *IEEE Winter Conference on Applications of Computer Vision*, pp. 2806 - 2815, 2021.
- [14] M. Véges, A. L'orincz, "Absolute Human Pose Estimation with Depth Prediction Network," *International Joint Conference on Neural Networks*, pp. 1 - 7, 2019.
- [15] G. Moon, J. Y. Chang, K. M. Lee, "V2V-PoseNet: Voxel-to-voxel Prediction Network for Accurate 3d Hand and Human Pose Estimation from a Single Depth Map," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5079 - 5088, 2018.
- [16] R. A. Guler, N. Neverova, I. Kokkinos, "Densepose: Dense Human Pose Estimation in the Wild," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7297 - 7306, 2018.
- [17] G. Rogez, P. Weinzaepfel, C. Schmid, "LCR-Net: Localization-classification-regression for Human Pose," *IEEE Conference on Computer Vision and Pattern Recognition*, pp.3433-3441, 2017.
- [18] OptiTrack Motive. Available online: <https://optitrack.com/software/motive/>
- [19] T. Maruyama, T. Ueshiba, M. Tada, H. Toda, Y. Endo, Y. Domae, Y. Nakabo, T. Mori, K. Suita, "Digital Twin-driven Human Robot Collaboration Using a Digital Human," *Sensors*, Vol. 21, No. 24 2021.
- [20] V. Weistroffer, F. Keith, A. Bisiaux, C. Andriot, A. Lasnier, "Using Physics-based Digital Twins and Extended Reality for the Safety and Ergonomics Evaluation of Cobotic Workstations," *Front. Virtual Real*, Vol. 3, pp.1-18, 2022.
- [21] J. D. Choi, M. Y. Kim, "A Sensor Fusion System with Thermal Infrared Camera and LiDAR for Autonomous Vehicles and Deep Learning Based Object Detection," *ICT Express*, Vol. 9, No. 2, pp. 222-227, 2023.
- [22] Y. Park, S. Yun, C. S. Won, K. Cho, K. Um, S. Sim., "Calibration Between Color Camera and 3d Lidar Instruments with a Polygonal Planar Board," *Sensors*, Vol. 14, No. 3, pp. 5333 - 5353, 2014.
- [23] H. Liu, H. Li, X. Liu, J. Luo, S. Xie, Y. Sun, "A Novel Method for Extrinsic Calibration of Multiple RGB-D Cameras Using Descriptor-based Patterns," *Sensors*, Vol. 19, No. 2, 2019.
- [24] S. Lee, J. Yoo, M. Park, J. Kim, S. Kwon, "Robust Extrinsic Calibration of Multiple RGB-D Cameras with Body Tracking and Feature Matching," *Sensors*, Vol. 21, No. 3, 2021.
- [25] N. Gelfand, L. Ikemoto, S. Rusinkiewicz, M. Levoy, "Geometrically Stable Sampling for the ICP Algorithm," *IEEE 3D Digital Imaging and Modeling*, 2003.
- [26] ISO/TS 15066:2016, "Robots and robotic devices - Collaborative robots", <https://www.iso.org/standard/62996.html>

**Ji Dong Choi (최 지 동)**

2022 School of Department of Future Automotive and IT Convergence from Kyungpook National University, Republic of Korea (M.S.)

2023~School of Electronic and Electrical Engineering from Kyungpook National University, Republic of Korea (Ph.D. Candidate)

## Career:

2022~2023 Researcher at the Korea Electronics Technology Institute

Field of Interests: Computer Vision, Pose Estimation, Object Detection, Autonomous Vehicle

Email: jdchoi@kitech.re.kr

**Min Young Kim (김 민 영)**

1996 Korea Advanced Institute of Science and Technology (B.S.)

1998 Korea Advanced Institute of Science and Technology (M.S.)

2004 Korea Advanced Institute of Science and Technology (Ph.D.)

2009~School of Electronic and Electrical Engineering from Kyungpook National University (Prof.)

## Career:

2004~2005 Senior Researcher at the Mirae Co., Korea.

2005~2009 Chief Researcher at the Kohyoung Co., Korea.

2013~Director at Research Center for Neurosurgical Robotic System, Kyungpook National University, Korea.

2014~2015 Associate Professor at the Department of Electrical and Computer Engineering and the School of Medicine, Johns Hopkins University

2016~Deputy Director at the KNU-LG Electronics Convergence Research Institute, Kyungpook National University, Korea.

Field of Interests: visual intelligence for robotic perception, recognition of autonomous unmanned ground and aerial vehicles

Email: minykim@knu.ac.kr

**Byeong Hak Kim (김 병 학)**

2015 School of Electronic Engineering from Kyungpook National University, Republic of Korea (M.S.)

2020 School of Electronic Engineering from Kyungpook National University, Republic of Korea (Ph.D.)

2020~Korea Institute of Industrial Technology (Senior Researcher) Career:

2007~2015 Engineer at the SAMSUNG THALES, Korea.

2016~2019 Engineer at the Hanwha Systems, Korea.

Field of Interests: IR and 3D Imaging Systems, Visual Object Tracking, ML/DL Object Detection, 3D Laser Radar, Counter Drone Systems, and Gesture Recognition

Email: bhkim81@kitech.re.kr