



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학석사 학위논문

디지털 트윈을 활용한 스마트 공간 내
인간-컴퓨터 간 상호작용 시스템 개발

Development of Human-Computer Interaction
System in Smart Space Using Digital Twin

울 산 대 학 교 대 학 원
전기전자컴퓨터공학과
김 준 명

디지털 트윈을 활용한 스마트 공간 내 인간-컴퓨터 간 상호작용 시스템 개발

지도교수 조강현

이 논문을 공학석사학위 논문으로 제출함

2023년 12 월

울산대학교 대학원
전기전자컴퓨터공학과
김준명

김준명의 공학석사학위 논문을 인준함

심사위원장	강	희	준	(인)
심사위원	조	강	현	(인)
심사위원	서	영	수	(인)

울 산 대 학 교 대 학 원
2023 년 12 월

[초록]

디지털 트윈을 활용한 스마트 공간 내 인간-컴퓨터 간 상호작용 시스템 개발

울산대학교 대학원
전기전자컴퓨터공학과
김 준 명

최근 스마트 홈에 관한 관심과 연구가 증가하고 있으며 일반 주택뿐만이 아닌 다양한 장소에서 활용되는 시도가 이어지고 있다. 스마트 홈을 구성하는 주요 기술 중 하나인 Human-Computer Interaction (HCI) 또한 여러 방향에서 연구가 진행되고 있다. 본 연구는 HCI를 구현할 때 주로 사용되는 방법들에 대한 문제점을 제기하고, 문제점에 대한 해결책으로 디지털 트윈과 인공지능 모델을 통해 HCI를 구축하는 방법을 제시한다. 실제 현실의 장소에 대한 규격을 측정하고, 측정한 데이터를 바탕으로 디지털 트윈 환경을 구축했다. 이후 3개의 카메라를 통해 사람의 2차원 관절 좌표를 추정, 각 관절의 좌표를 통해서 3차원 관절 좌표를 설정하여 사람의 3차원 상태를 디지털 트윈 내에 표현, 자세를 통하여 상호작용을 발생시킨다. 위와 같은 순서를 통해 HCI 시스템을 구축하고, 제안한 시스템의 유효성을 증명하기 위해 시뮬레이션 환경에서 5개의 상호작용 객체에 대해서 상호작용의 성공률에 대해 실험을 수행하였고, 최종적으로 82.7%의 성공률을 확인하였다.

[Abstract]

Development of Human-Computer Interaction System in Smart Space Using Digital Twin

Junmyeong Kim

School of Electrical Engineering,

The Graduate School,

University of Ulsan

Supervised by Prof. Kanghyun Jo

Recently, there has been an increasing interest and research in smart homes, with experiments taking place in various locations beyond just regular homes. Human-Computer Interaction (HCI), a key technology in constructing smart homes, is also being researched from various angles. This study raises issues regarding the methods commonly used in implementing HCI and proposes a solution to these problems by building HCI through digital twins and artificial intelligence models. Real-world spaces were measured to create a digital twin environment based on the collected data. Subsequently, 2D joint coordinates of a person were estimated using three cameras, and 3D joint coordinates were set using these joint coordinates to represent the three-dimensional state of a person within the digital twin, generating interactions through posture. Following this sequence, an HCI system was established, and experiments on interaction with five interactive objects were conducted in a simulation environment to demonstrate the effectiveness of the proposed system. Ultimately, when measuring interactions using three cameras, an accuracy of 82.7% was achieved.

목차

국문요약	I
영문요약	II
목차	III
그림 목차	IV
표 목차	V
1. 서론	1
1.1. 논문의 구성	1
1.2. 연구 배경	2
1.2.1. 스마트 홈의 정의	2
1.2.2. 스마트 홈 시장 동향	4
1.2.2.1. 국내 스마트 홈 시장 동향	4
1.2.2.2. 해외 스마트 홈 시장 동향	5
1.2.3. 사람-컴퓨터 간 상호작용	6
1.2.3.1. 물리적인 디바이스를 통한 HCI 구현	7
1.2.3.2. AR Glass를 통한 HCI 구현	8
1.2.3.2. AI 스피커를 통한 HCI 구현	8
1.2. 연구 목표 및 내용	9
2. 선행 연구	10
2.1. 카메라 기하학	10
2.1.1. 스테레오 비전	11

2.2. 사람 자세 추정	12
2.1.1. Top-down 방법	12
2.1.2. Bottom-up 방법	12
2.3. 디지털 트윈	13
3. 디지털 트윈 기반 HCI 구현	14
3.1. 디지털 트윈 구축	14
3.2. 2차원 관절위치 검출	16
3.2.1. 사람 검출	16
3.2.2. 이미지 내 관절 위치 검출	16
3.3. 3차원 관절위치 추정 및 상호작용 구현	17
3.3.1. 3차원 관절 위치 추정	17
3.3.2. 상호 작용 구현	21
4. 실험	24
4.1. 실험 환경	24
4.1.1. 시뮬레이션 환경	24
4.1.2. 상호작용 설정	24
4.1.3. 데이터 통신 및 인공지능 모델	25
4.1.4. 카메라와 PC 환경	26
4.1.5. 카메라 파라미터	26
4.2. 실험 결과	28
4.2.1. 시뮬레이션 환경	28
4.2.2. 상호작용 성능	29
5. 결론	30
참고문헌	31

그림 목차

그림 1.1 스마트 홈	2
그림 1.2 국내외에서 제공하는 스마트 홈 서비스	3
그림 1.3 사람과 컴퓨터 간 상호작용	6
그림 1.4 물리적인 디바이스를 통한 HCI 구현	7
그림 1.5 AR Glass를 통한 HCI 구현	8
그림 1.6 전체 프로세스	9
그림 2.1 핀홀 카메라 모델	10
그림 2.2 스테레오 비전을 활용한 물체까지의 거리 추정	11
그림 2.3 Top-down 방법	12
그림 2.4 Bottom-up 방법	12
그림 3.1 다양한 시점에서 본 디지털 트윈 환경	14
그림 3.2 2개의 영상을 통한 3차원 투영	17
그림 3.3 3개의 카메라를 사용했을 때 3차원 좌표	19
그림 3.4 3차원 좌표계에 투영된 관절	20
그림 3.5 상호작용 과정	21
그림 3.6 디지털 트윈에서의 상호작용 발생 과정	23
그림 4.1 Top-view 시점에서 본 실험 환경	24
그림 4.2 Firebase	25
그림 4.3 G-GOON GPRO-QHD800 4K	26
그림 4.4 3개의 카메라 시점에서 실제 환경과 디지털 트윈	27
그림 4.5 투영된 3차원 관절 위치	28
그림 4.6 관절의 위치를 통한 캐릭터 투영	28

표 목차

표 1.1 국내 스마트 홈 시장 규모 추이	4
표 1.2 세계 스마트 홈 시장 규모 추이	5
표 1.3 세계 스마트홈 설치 가구 수	5
표 4.1 추정하는 관절의 종류	25
표 4.2 디지털 트윈 PC와 자세추정 PC 성능	26
표 4.3 카메라 파라미터	27
표 4.4 카메라 2개를 사용했을 때 상호작용 성능	29
표 4.5 카메라 3개를 사용했을 때 상호작용 성능	29

I. 서론

1.1. 논문의 구성

본 논문은 총 5장으로 구성되어 있으며 다음과 같은 순서로 내용을 소개한다.

먼저 1장에서는 스마트 홈의 정의 및 시장 규모를 통해 현재 국내 및 국제적으로 스마트 홈의 중요성이 커지고 있다는 내용을 전달한다. 이후 현재 존재하는 HCI 구현 방법과 문제점을 제기하여 본 연구의 필요성을 설명한다. 이후 연구 내용과 목표에 대해 간략하게 전달한다.

2장에서는 해당 연구를 위해 사용된 카메라 기하학, 사람의 자세 추정, 그리고 디지털 트윈에 대한 개념을 설명한다.

3장에서는 본 연구에서 제안한 HCI 시스템을 구현하는 방법에 대해 구체적인 방법을 설명한다.

4장은 3장에서 설명한 방법에 대한 성능 평가 및 실험 환경에 대한 자세한 정보를 전달한다.

5장에서는 4장의 실험 결과를 통해 본 논문의 효용성, 개선 가능성을 설명하고 전체적인 내용을 정리한다.

1.2. 연구 배경

1.2.1. 스마트 홈의 정의

스마트 홈 (Smart Home)은 인공지능과 IoT (Internet of Things) 기술을 활용하여 건물 내부의 다양한 기기와 시스템을 연결하여 자동화 하는 것을 말한다. 이를 통해 편리성, 안정성, 에너지 효율성, 의료 서비스 등 여러 가지 장점을 사용자에게 제공하여 거주자의 삶의 질을 향상시키는 것이 가능하다. 스마트 홈이란 용어는 일반적인 주택에 제한되지 않으며 병원, 학교 등과 같은 공간에서도 적용되어지며 공간의 목적에 맞는 기능을 사용자에게 추가적으로 제공하기도 한다.

스마트 홈 (Smart Home)은 인공지능과 IoT (Internet of Things) 기술을 활용하여 건물 내부의 다양한 기기와 시스템을 연결하여 자동화하는 것을 말한다. 이를 통해 편리성, 안정성, 에너지 효율성, 의료 서비스 등 여러 가지 장점을 사용자에게 제공하여 거주자의 삶의 질을 향상하는 것이 가능하다. 스마트 홈이란 용어는 일반적인 주택에 제한되지 않으며 병원, 학교 등과 같은 공간에서도 적용되며 공간의 목적에 맞는 기능을 사용자에게 추가로 제공하기도 한다.

스마트 홈의 개념이 처음 등장한 것은 20세기 중반으로, 스마트 홈의 전신이 되는 “홈 오토메이션(Home Automation)”이란 용어의 줄임말인 “홈 오토(Home Auto)”가 등장하면서이다. 이 단어는 스위치로 냉난방 및 각종 전자제품을 제어 및 전자 우편과 컴퓨터를 사용한 재택근무 등의 집안에서 일어나는 일을 제어 및 자동화하는 기술을 포괄적으로 표현하는 단어이다. “홈 오토”의 등장과 함께



그림 1.1 스마트 홈

현재는 국내외 다양한 대기업 혹은 중견 기업에서 스마트홈 서비스를 제공하고 있는데, 삼성에서 제공하는 “SmartThings” 애플리케이션을 통한 일반 주택 내 가구 또는 전자 기기 제어, 애플에서 아이폰의 음성 AI인 시리를 통한 스마트홈 기기 제어 플랫폼 “애플 홈킷(Apple HomeKit)”, 애플과 같이 KT에서 서비스하는 음성 AI인 Genie를 통해 스마트홈 시스템을 구현한 “GiGa Genie” 등이 있다.



(a) 삼성의 SmartThings (b) 애플의 Apple HomeKit (c) KT의 GiGa Genie
그림 1.2 국내외에서 제공하는 스마트 홈 서비스

1.2.2. 스마트 홈 시장 동향

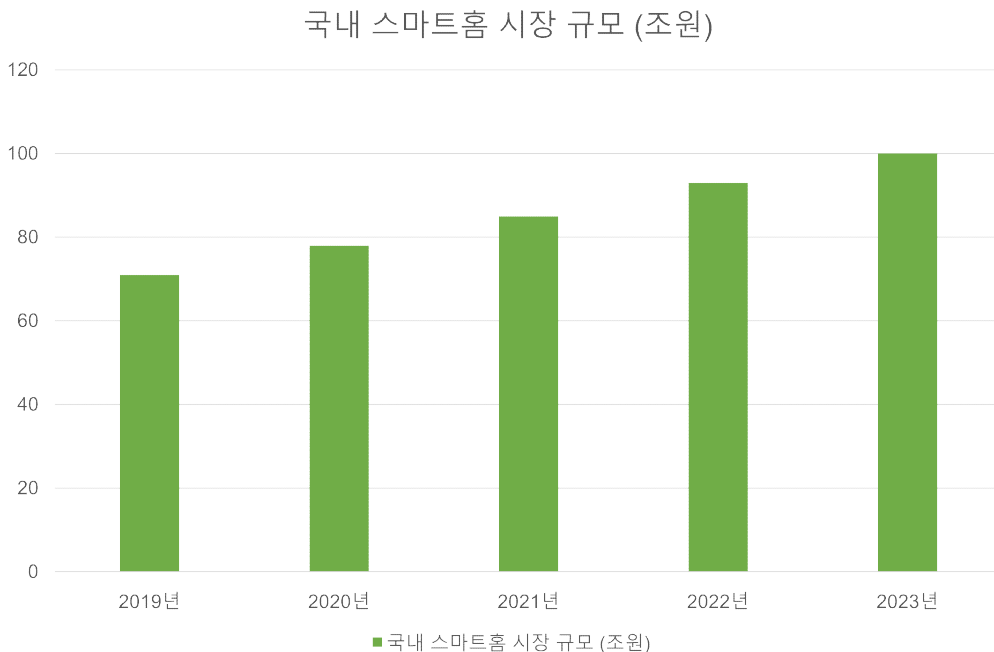
스마트 홈은 2019년에서 시작된 COVID-19의 확산으로 시작된 팬데믹(Pandemic) 사태로 인해 야외 활동의 규제와 함께 국내, 국외 모두에서 매년 큰 폭으로 시장의 규모, 스마트 홈 시스템 설치 가구 수 등이 증가하고 있다. 해당 항에서는 그에 관한 내용을 다룬다.

1.2.2.1 국내 스마트 홈 시장 동향

표 1.1은 “한국 AI 스마트홈 산업협회”에서 조사한 2019년부터 2023년까지 변화한 국내의 스마트 홈 시장의 규모를 나타낸다. 2019년부터 2023년까지 약 10퍼센트가량 꾸준히 증가한 것을 확인할 수 있다. 또한, 2019년과 2023년을 비교했을 때 30조가량 시장 규모가 증가한 것을 확인할 수 있다. 표 1.1과 같이 스마트 홈 시장의 규모가 커지면서 아파트 및 주택에서도 음성인식, 지문인식, 안면인식 등의 기술을 도입하여 스마트 홈 관련 기술을 도입하기 위한 시도를 거듭하고 있다.

표 1.1 국내 스마트 홈 시장 규모 추이

	2019년	2020년	2021년	2022년	2023년
규모(조원)	70조	78조	85조	93조	100조
규모(억)	9698억	2837억	7048억	728억	4455억



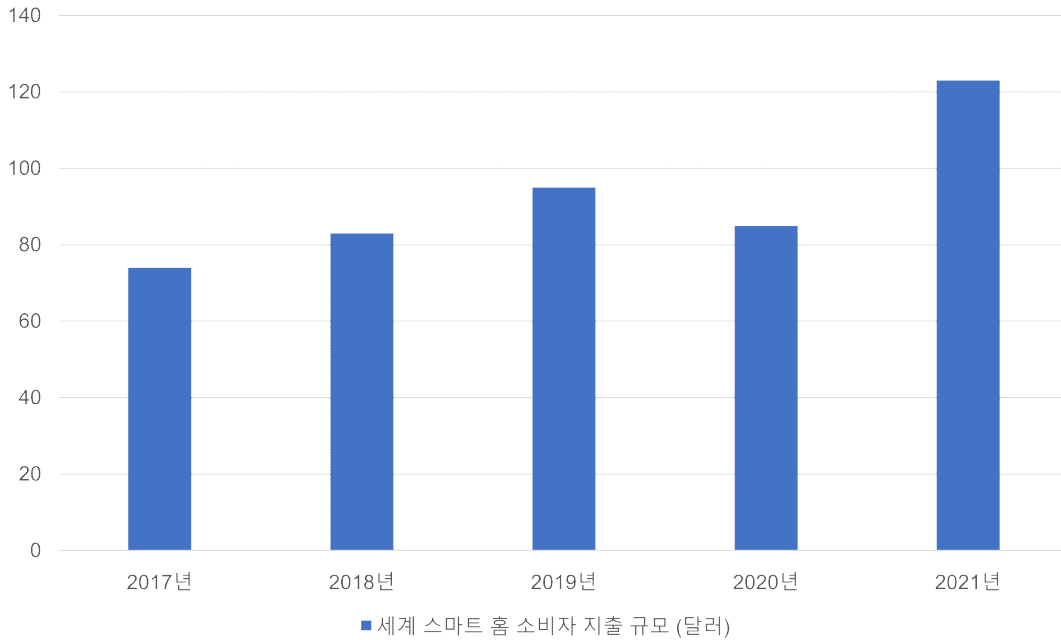
1.2.2.2. 국외 스마트홈 시장 규모

세계적으로도 스마트 홈 시장의 규모가 매년 증가하는 것을 확인할 수 있는데, 표 1.2는 미국 시장조사업체인 “스트래티지 애널리틱스(Strategy Analytics, SA)”가 조사한 “2021년 세계 스마트 홈 전망” 보고서에서 스마트 홈에 대한 평균 소비자 지출 금액을 나타낸다. 해당 표에 2017년 이후 2020년을 제외하곤 2021년까지 꾸준히 소비자 지출 금액이 증가하는 것을 보여주며 2021년에는 123달러까지 증가하는 것을 보여준다.

표 1.2 세계 스마트 홈 소비자 지출 규모

	2017년	2018년	2019년	2020년	2021년
규모(달러)	74	83	95	85	123

세계 스마트 홈 소비자 지출 규모 (달러)



스마트 홈 설치 가구 또한 세계적으로 매년 증가하는 추세를 보인다. 2017년에는 1억 8,600만 가구에서 2021년에는 약 3억 가구가량으로 증가했다. 표 1.3은 스마트 홈 가구 설치 수에 대한 2017년부터 2021년까지의 통계를 나타낸다.

표 1.3 세계 스마트 홈 설치 가구 수

	2017년	2018년	2019년	2020년	2021년
규모(백만)	186	219	254	276	306

1.2.3. 사람-컴퓨터 간 상호작용 (Human-Computer Interaction, HCI)

스마트 홈은 인공지능, 통신, IoT 기술, 자동화 등 다양한 기술에 의해 구성된다. 그 중의 가장 최우선으로 되는 것은 사람과 컴퓨터 간의 상호작용(Human-Computer Interaction, HCI)을 구현하는 것이다. HCI는 컴퓨터와 사람 간의 상호작용을 뜻하기도 하지만, 컴퓨터 기술과 사용자 간의 상호작용을 연구하고 설계하여 학문 분야를 나타내기도 한다. 이를 통해, 블루투스 스피커가 사용자의 음성을 인식해 창문을 닫게 하고, 리모컨을 통해 전등 또는 TV의 화면을 제어하는 모든 것이 HCI에 포함된다. 그림 1.3은 HCI의 개념에 대해 나타낸다. 본 항에서는 HCI를 구현할 때 일반적으로 어떤 방법이 사용되는지, 사용하는 방법에 대한 문 제점은 어떤 것이 있는지 다룬다.



그림 1.3 사람과 컴퓨터 간 상호작용 (Human-Computer Interaction, HCI)

1.2.3.1. 물리적인 디바이스를 통한 HCI 구현

HCI를 구현할 때 일반적으로 사용되는 방법은 키보드, 스마트폰 등과 같은 물리적인 장치를 사용해서 구현하는 것이 가장 자주 사용되는 접근 방식이다. 넓은 의미로 봤을 때, Remote Controller를 사용해서 TV의 화면을 바꾸고, 스마트폰을 통해서 블루투스 스위치를 컨트롤해 LED 조명을 제어하는 것 또한 이 방법에 해당한다. 이런 방법은 사용자의 지시 명령을 직접적으로 HCI 시스템에 전달하기 때문에 비교적 구현하기 쉬운 장점이 있다. 하지만 단점 또한 존재하는데, 먼저 물리적인 장치가 사용자가 보유하고 있지 않으면 상호작용을 만들 수 없다는 단점이 존재한다. 이러한 장비에 대한 의존성은 사용자가 장치를 물리적으로 제어해야 하는 불편함이 존재한다. 또한, 사용 방법에 대한 직관성이 떨어져 노인과 아동이 사용할 때 사용의 어려움을 일으킬 수 있다는 단점이 있다.



그림 1.4 물리적인 디바이스를 통한 HCI 구현

1.2.3.2. AR Glass를 통한 HCI 구현

최근 메타버스와 관련된 증강 현실(Augmented Reality, AR), 가상 현실(Virtual Reality, VR), 확장 현실(Extended Reality, XR), 혼합 현실(Mixed Reality, MR) 기술의 발전으로 AR Glass를 통해 HCI를 구현하는 방법이 연구되고 있다. AR Glass의 경우 사람의 눈과 손을 통해 AR Glass의 디스플레이에 표시된 인터페이스와 상호작용하여 HCI를 동작시킨다. 이는 다른 물리적인 기기를 통해 HCI 구현하는 방법과 달리 사용자의 손과 눈을 통해 직접적으로 상호작용할 수 있게 하여 처음 접한 사용자들도 비교적 쉽게 사용할 수 있게 됐다. 하지만 이 경우 장비의 높은 가격, 제한된 디스플레이 각도, 낮은 해상도, 부족한 배터리 수명, 사용자가 오래 사용하기 어려운 무게와 같은 문제가 있다. 위와 같은 문제로 인해 사용자층은 직관적인 사용 방법에 비해 제약이 있다.

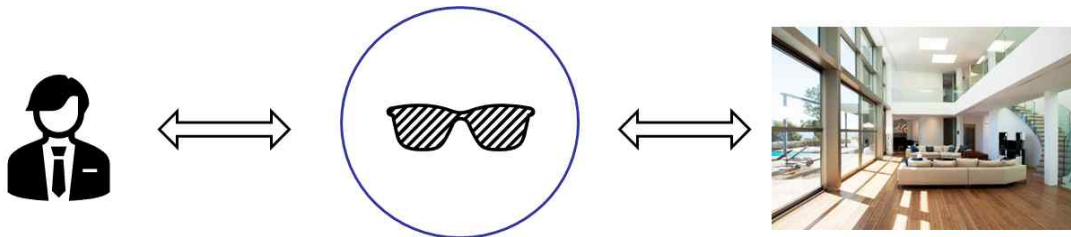


그림 1.5 AR Glass를 통한 HCI 구현

1.2.3.3. AI 스피커를 통한 HCI 구현

다른 방법으로는 AI 스피커를 활용해 HCI를 구현하는 방법 또한 존재한다. 이 경우 음성 인식 AI를 사용자의 음성 명령을 통해 컴퓨터와 상호작용을 발생시키는 방식으로 구현된다. 이를 통해 별도의 착용 및 물리적인 장비가 없이 사용자가 사용할 수 있다는 장점이 있다. 하지만 음성 인식 AI 능력에 대한 의존도가 커 잘못된 인식을 발생시킬 수 있는 문제가 있다. 또한 AI 스피커는 항상 대기 상태에서 사용자의 음성을 수집하고 행동을 처리한다. 이에 따라 개인 정보와 사생활 침해의 문제가 발생할 수 있다.

1.3. 연구 목표 및 내용

본 논문에서는 컴퓨터 비전 기술과 공간 내부에 존재하는 복수의 카메라, 실제 환경을 바탕으로 만들어진 디지털 트윈을 활용하여 스마트 홈 시스템을 위한 HCI를 구현 및 개발을 목표로 한다. 구현을 위해 3차원 좌표가 표현되기 위한 현실과 유사한 디지털 트윈 구축, 사람 검출, 관절 검출을 통해 2차원 좌표를 3차원 좌표로 변환한다. 이후 검출한 관절의 3차원 좌표를 사용하여 문과 냉장고와 같은 다양한 가구와 상호 작용 구현을 수행한다. 각 단계를 구현함으로써 별도의 착용 장비 및 물리적인 기기의 제약이 없는 HCI 시스템을 구현한다. 그림 1.6은 해당 연구 내용의 프로세스를 간략하게 묘사하고 있다.

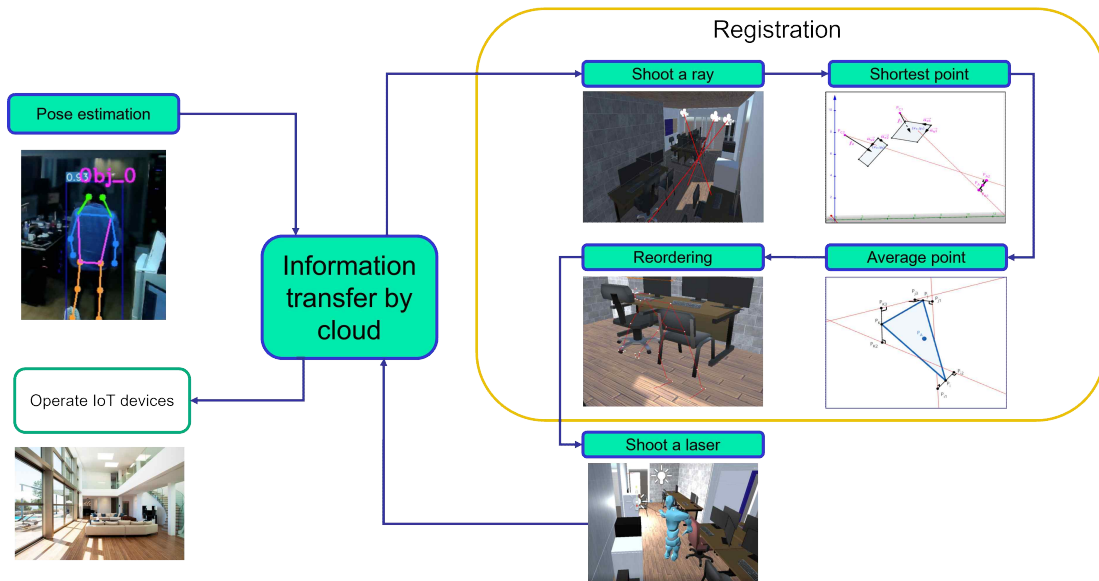


그림 1.6 전체 프로세스

먼저, 컴퓨터 비전을 기반으로 하는 딥러닝 알고리즘 중 하나인 자세 추정 방법을 통해 사람의 관절을 추출한다. 추출된 관절 정보는 클라우드 서비스 저장소 또는 직접적으로 다른 노드로 전달되고 그 노드에서는 정합 과정을 통해 여러 카메라로부터의 관절을 3차원 디지털 트윈 환경상에 매핑한다. 정합 시에는 각 카메라로부터 하나의 점으로 직선을 긋고 두 직선 사이의 거리를 결정하는 두 점을 계산한 후, 평균 지점을 계산한다. 이렇게 3차원 디지털 트윈을 구성하는 뼈대가 완성되면 뼈대 위에 사람의 형상을 입히는 리깅을 수행해 실제 사람처럼 보이도록 알고리즘을 구성한다. 사람의 팔이 일정 각도 이상 벌어질 경우, 레이저를 발사하여 특정 객체에 맞으면 그 객체가 미리 정해진 알고리즘에 따라 동작하도록 한다. 실제 IoT 기기를 사용한다면 객체와 레이저의 충돌 정보는 클라우드 서비스를 통해 물리 기기로 전달되어 동작 시킨다.

2. 선행 연구

2.1 카메라 기하학

카메라 기하학은 컴퓨터 비전 및 컴퓨터 그래픽스 분야에서 핵심적인 개념 중 하나로, 디지털 이미지와 3D 공간 간의 관계를 정의하고 분석하는 데 사용된다. 이 개념은 카메라와 3차원 공간 사이의 상호작용을 이해하고 모델링하기 위한 중요한 도구로서, 이미지 처리, 로보틱스 (Robotics), 가상 현실 및 기타 다양한 응용 분야에서 중요한 역할 수행한다. [1]

카메라 기하학의 핵심 요소는 내부 파라미터와 외부 파라미터로 나눌 수 있다. 내부 파라미터는 카메라의 자체의 특성을 설명하며 초점 거리(Focal Length), 주점(Center Point), 비대칭 계수(Skew Coefficient) 등을 포함한다. 이러한 내부 파라미터는 카메라 렌즈와 이미지 센서의 특성을 표현한다. 외부 파라미터는 카메라의 물리적 실제 위치와 방향을 나타낸다. 위에서 말한 카메라 파라미터를 사용하여 카메라의 기하학적 분석이 이루어진다.

카메라 기하학은 핀홀 카메라 모델을 기반으로 하는데, 핀홀 카메라 모델은 카메라의 렌즈가 무한히 얇다고 가정하고, 빛이 카메라의 주점을 통과하여 이미지 센서 또는 필름에 도달하는 방식의 카메라를 뜻한다. 아래 그림 . 핀홀 카메라가 3차원 객체를 2차원에 영상 좌표계에 어떻게 투영되는지 나타낸다.

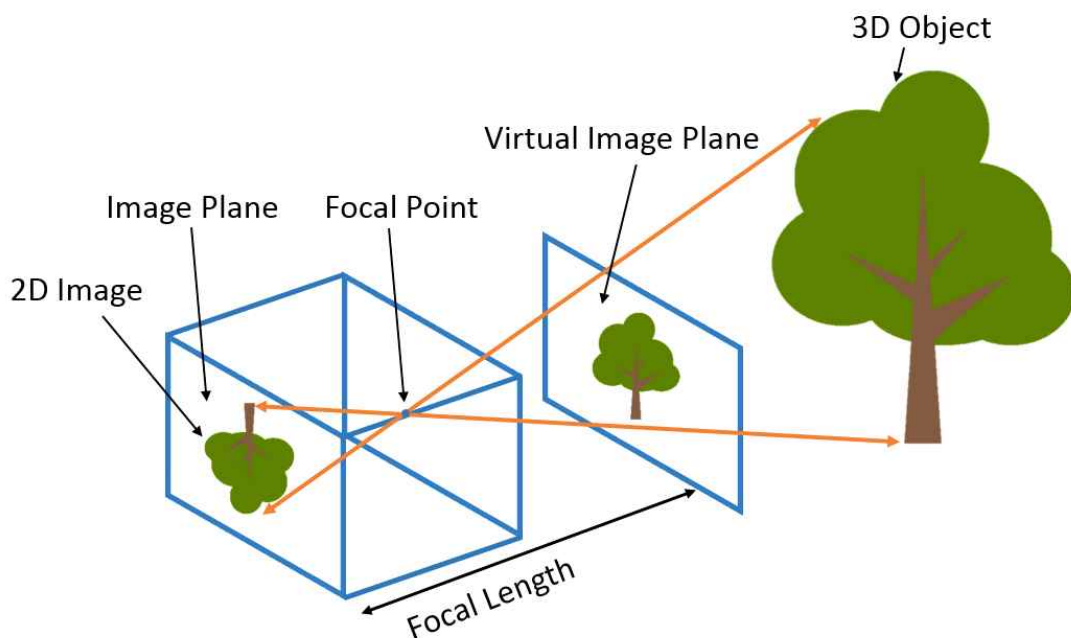


그림 2.1 핀홀 카메라 모델

2.1.1. 스테레오 비전

앞서 카메라 기하학과 함께 핀홀 카메라 모델에 대해 설명했다. 스테레오 비전 기술은 이러한 카메라 기하학을 통해 두 개의 카메라로부터 물체의 3차원 위치를 결정할 수 있게 한다. 그림 2.2는 두 개의 카메라로부터 물체까지의 거리를 추정하는 간단한 원리에 대한 그림을 나타낸다. 실제 3차원 좌표계에서는 그림과 달리 y 좌표에 대해 고려할 필요가 있지만 스테레오 비전의 원리를 간단하게 이해하기 위해 2차원상에서의 카메라 좌표를 고려한다.

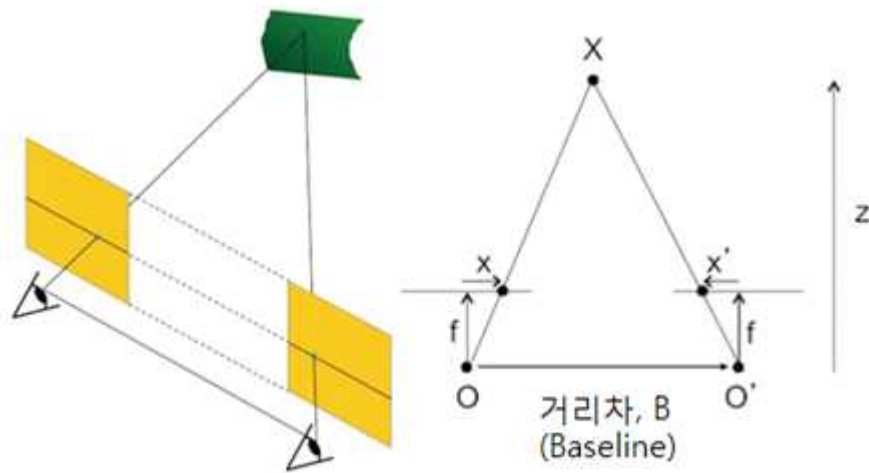


그림 2.2 스테레오 비전을 활용한 물체까지의 거리 추정

그림에서 f 는 카메라의 초점거리, x 와 x' 는 두 개의 카메라로부터 매칭된 각 점, X 는 물체의 위치, Z 는 물체까지의 거리, B 는 두 카메라 사이의 거리를 나타낸다. x 와 x' 의 원점은 각 이미지의 중심이고 오른쪽을 (+) 방향이라고 가정했다.

$$\frac{Z-f}{Z} = \frac{B-(x-x')}{B} \quad (1)$$

$$d = x - x' \quad (2)$$

$$Z = \frac{Bf}{d} \quad (3)$$

시차 d 는 두 이미지에서 매칭되는 한 점에 대한 픽셀 좌표의 차이이다. 매칭되는 두 점을 알고 있다면 시차를 식 (2)와 같이 계산할 수 있으므로 식 (3)을 통해 물체까지의 거리 Z 를 계산할 수 있다. 수식으로부터 시차가 커지면 물체가 가까이 있음을 의미한다는 것을 알 수 있다

2.2. 사람 자세 추정

사람의 자세 추정은 하나의 이미지 혹은 비디오에서 사람의 관절과 같은 Key point를 탐지하고 추정하는 분야를 말한다. 이런 사람의 자세 추정 방법은 크게 Top-down 방법과 Bottom-up 방법으로 구분된다.

2.2.1. Top-down 방법

먼저 Top-down 방법은 YOLO(You Only Look Once) [2], Faster-RCNN [3] 와 같은 객체 검출 모델을 활용하여 영상 속 존재하는 모든 사람의 영역을 개별적으로 추정하고, 추정한 영역 안에서 사람의 자세 추정하는 방법을 말한다. 해당 방법은 사람이 있는 위치를 별도의 검출 모델을 사용해 추정하는 만큼 우수한 성능을 보이지만, 각 사람의 영역에 대해 자세를 추정하는 만큼 사람의 수에 따라 연산 속도가 느려지는 특징을 보여준다.

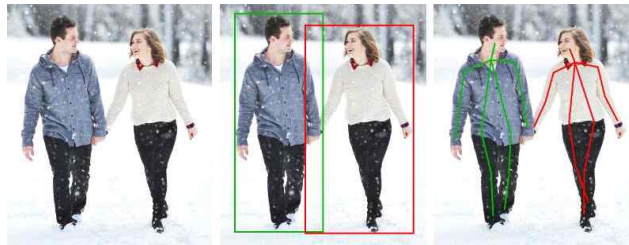


그림 2.3 Top-down 방법

2.2.2. Bottom-up 방법

반면 Bottom-up 방법의 경우 영상 속의 모든 관절 정보를 추정하고 각 관절이 어느 집합에 속하는지를 추정해 사람의 자세를 추정한다. 이 방법은 별도의 객체 검출 모델을 사용하지 않고 직접적으로 관절의 위치 정보를 추정하기 때문에 검출 성능이 떨어진다는 단점이 있지만 비교적 사람의 수에 대한 영향이 Top-down 방법에 비해 적고, 연산량이 적다는 장점이 있다. 본 연구에서는 정확한 사람의 관절 정보를 취득하기 위해 비교적 높은 성능을 보이는 Top-down 방법을 채택했다. Bottom-up에 대표적인 방법으로는 OpenPose [4] 가 있다



그림 2.4 Bottom-up 방법

2.3. 디지털 트윈

디지털 트윈이란 현실 세계의 기계나 장비, 사물 등을 컴퓨터 속 가상 세계에서 구현하는 것을 의미한다. [5] 처음 디지털 트윈이 구현된 것은 1960년 NASA Apollo 13으로, 기술자들이 지상에서 우주 공간에서 진행되는 구조 임무에 대해 가능한 방법을 검증하기 위해 만들어졌다. 이후, 디지털 트윈이란 용어를 처음으로 정의한 것은 Michael Grieves가 2003년에 발표한 제품 수명 주기 관리에 관한 책에서 다양한 형태로 제안되어 등장했다. [6]

최근 디지털 트윈 기술은 마이크로프로세서와 반도체 기술의 발달로 인해 하나의 컴퓨터로 연산할 수 있는 양이 늘어나고 실시간성이 증가하며 발전하였다. 산업에서는 관제 시스템, 감시 시스템 등에 많이 사용된다. 대표적으로는 선박 내의 수많은 시스템을 디지털 트윈화 하여 관제소에서 점검하며 자율운항하는 기술이 있다. 그뿐만 아니라 디지털 트윈을 통해 실제 도로에서 실험하면 수없이 많이 생기는 피해와 손실을 줄일 수 있다. 실제 환경이랑 똑같은 환경을 구축해두고 그 환경에서 자율주행 자동차의 알고리즘을 실험한다면 사고로 인해 발생하는 인명피해, 경제적 손실 등을 줄일 수 있다.

디지털 트윈 환경을 구축하는 방법은 여러 가지가 있을 수 있는데, 첫 번째는 수동으로 모든 환경을 구축하는 것이다. 이러한 방법은 실제 환경의 길이 등을 측정해 디지털 트윈 상에 반영하는 것이다. 예를 들어 실외와 같은 디지털 트윈 환경을 구축할 때는 이러한 방법이 적합하지 않지만, 실내의 공간을 구축할 때는 매우 정밀하게 구축할 수 있다는 장점이 있다. 두 번째는 라이다나 카메라와 같은 센서들을 통해 실제 환경을 스캔하는 방법이다. 스캔된 데이터를 가공하여 디지털 트윈 환경을 구축하면 반자동으로 환경을 구축하기 때문에 매우 빠른 시간 내에 광범위한 범위의 환경을 구축할 수 있다. 이 방법은 첫 번째 방법보다 효율적이지만 디지털 트윈을 구축하는 알고리즘에 따라 부정확하다는 단점이 있다. 본 논문에서는 첫 번째 방법을 통해 디지털 트윈 시스템을 스마트 홈이나 최근 주목받고 있는 스마트 공간 시스템에 적용할 수 있도록 구축하고 여러 실험을 진행한다.

3. 디지털트윈 기반 HCI 구현

3.1. 디지털 트윈 구축

본 연구는 실제 사물과 비슷한 디지털 트윈 월드를 구현하기 위해 3D 게임 엔진인 Unity를 활용했다. 실제 방의 규격과 가구의 위치를 직접 측정, 측정한 값을 바탕으로 방과 유사한 3차원 환경을 구축했다. 아래 그림 3.1은 해당 환경을 바탕으로 만들어진 디지털 트윈을 각 여러 시점에 촬영한 모습을 나타낸다. 3.3절의 방법을 통해 검출된 사람의 관절 3차원 좌표를 디지털 트윈 환경에 위치시키고, 매시간 새롭게 갱신되며 사람의 관절 위치를 디지털 트윈 내에 반영시킨다. 이 정보를 통해 디지털 트윈에서 HCI를 발생시키고, 이를 통해 현실의 디지털 트윈에 정보가 반영되게 한다.





그림 3.1 다양한 시점에서 본 디지털트윈 환경

3.2. 2차원 관절위치 검출

2차원의 관절 위치 검출은 총 3단계의 구성으로 이루어진다. 먼저 카메라 영상 속에서 사람을 검출한다. 검출된 바운딩박스 영역에 한해서 자세 추정을 수행한다. 이 방법을 설치된 복수의 카메라에 대해서 각각 수행함에 따라 여러 개의 관절 위치 정보의 집합을 얻을 수 있다. 아래 그림 3.2 은 하나의 카메라에서 일어나는 프로세스를 도식화한 것이다.

3.2.1. 사람 검출

사람의 검출을 위해서 Convolutional Neural Network (CNN) 기반의 객체 검출 모델을 사용했다. 객체 검출 네트워크는 크게 1단계 검출기(One-Stage Detector), 2단계 검출기로 구분된다. 먼저 1단계 검출기는 객체의 위치 추정과 식별이 동시에 이루어진다. 반면, 2단계 검출기의 경우 객체의 위치 추정과 식별이 각 단계에서 수행된다. 과거 연구에서는 1단계 검출기 검출기가 위치 추정과 객체 식별이 동시에 이루어져 속도는 빠르나 성능이 2단계 검출기에 비해 비교적 떨어진다는 연구가 있었지만 최근 객체 검출 모델의 연구가 진행됨에 따라 1단계 검출기의 성능이 2단계 검출기에 비슷하거나 능가하는 성능을 보이는 경우가 있어 1단계 검출기의 성능이 많이 쓰이는 추세이다. 이에 따라 본 연구에서는 1단계 검출기 중 하나를 사용하여 학습했다.

3.2.2 이미지 내 관절 위치 검출

사용자로 식별된 대상에 대해서 영상 속 관절의 2차원 좌표를 검출하기 위해 인공지능 자세 추정 네트워크를 사용했다. 사람의 코, 어깨, 골반 등 총 17개의 관절을 2차원 좌표를 검출한다. 검출된 결과는 실시간으로 사전에 구축한 데이터베이스에 저장된다.

3.3. 3차원 관절 위치 추정 및 상호작용 구현

3.3.1 3차원 관절 위치 추정

3개의 카메라에서 검출된 관절의 2차원 좌표 집합을 바탕으로 3차원 좌표로 변환한다. 변환을 수행할 때 본 연구에서는 카메라 기하학을 사용하는 방법을 채택했다. 각 카메라에서 포착된 관절의 위치는 이미지 좌표계에서 하나의 점으로 표현된다. 해당 점은 3차원 공간의 하나의 점으로 표현할 수 있고, 3개의 카메라에서 추출한 관절 정보를 모두 투영, 조합함으로써 디지털 트윈 내의 좌표계에 표현되는 하나의 점의 위치를 선정할 수 있게 된다.

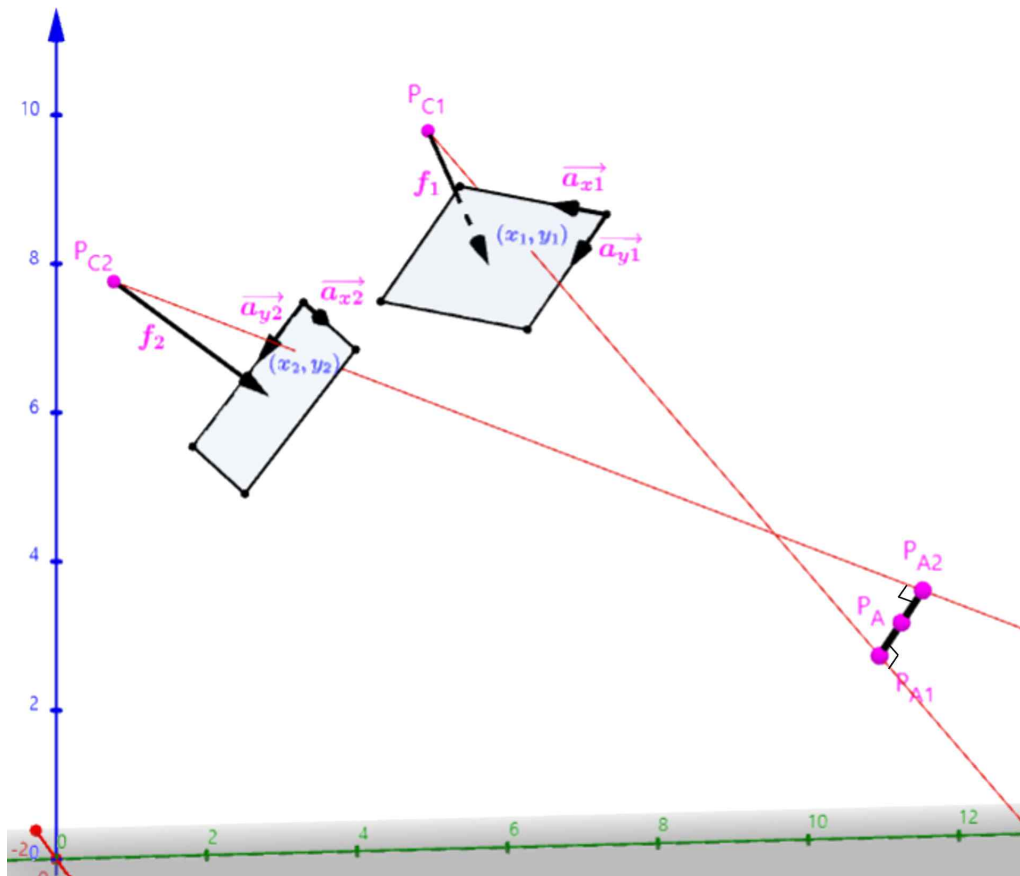


그림 3.2 2개의 영상을 통한 3차원 투영

그림 3.2는 02개의 이미지가 주어졌을 때 각 점이 3차원 공간 속에서 투영되는 방법을 보여 준다. 파란색의 x^i 와 y^i 는 i 번 카메라에서 포착된 이미지 좌표계에서의 점의 위치를 나타낸다. 해당 점의 단위를 픽셀에서 미터 단위로 바꿔주기 위해 카메라의 총해상도와 카메라의 센서 사이즈 정보를 활용하여 식 (1) 과 식 (2) 을 사용하여 변환시킨다. x_{im} 과 y_{im} 은 미터 단위로 변환된 점의 좌표를 나타낸다.

$$x_{im}^i = x^i / resolution * SensorSize \quad (1)$$

$$y_{im}^i = y^i / resolution * SensorSize \quad (2)$$

좌표계를 미터 단위로, 변환한 후 현실 좌표계로 좌표를 변환하기 위해 식 (2)를 사용한다. 그림 3.4에서 보라색 색상은 모두 월드 좌표 시스템을 기반이란 것을 나타낸다. P_C^i 는 i 번째 카메라의 위치 좌표, f^i 는 초점 거리 벡터, w^i 와 h^i 는 카메라 센서의 수평, 수직 길이를 나타낸다. P_{im}^i 은 현실 좌표계로 x_{im} 과 y_{im} 변환한 좌표이다.

$$P_{im}^i = P_C^i + f^i + (x_{im} - w/2)\vec{a}_x^i + (y_{im} - h/2)\vec{a}_y^i \quad (2)$$

이 경우 i 번째 카메라 위치 P_C^i 에서 바라본 P_{im}^i 에 대한 거리 벡터 D_{imC}^i 는 식 (3)과 같이 계산된다. D_{imC}^i 를 무한히 확장함으로써 그림 3.4에서 그려진 붉은 선과 같은 벡터를 얻을 수 있다.

$$D_{imC}^i = P_{im}^i - P_C^i \quad (3)$$

P_A^1 과 P_A^2 는 각 D_{imC}^i 를 확장한 붉은 선 위의 점이라고 했을 때, 각 점의 위치는 식 (4)과 식 (5)의 형태로 표현된다.

$$P_A^1 = P_C^1 + t_1 D_{imC}^1 \quad (4)$$

$$P_A^2 = P_C^2 + t_2 D_{imC}^2 \quad (5)$$

P_A^1 과 P_A^2 에 대한 거리 벡터와 D_{imC}^i 의 관계가 수직일 때 두 점간의 거리가 최소한이 된다면 내적했을 때 값이 0인 성질을 이용해 식 (6)과 (7)의 두 개의 방정식의 해를 푸는 것으로 P_A^1 과 P_A^2 값을 계산할 수 있다.

$$(P_A^1 - P_A^2) \cdot D_{imC}^1 = 0 \quad (6)$$

$$(P_A^1 - P_A^2) \cdot D_{imC}^2 = 0 \quad (7)$$

$$P_A = \frac{(P_A^1 + P_A^2)}{2} \quad (8)$$

두 개의 카메라에서만 검출된 경우, 식 (8)과 같이 두 점의 평균을 통해서 디지털 트윈에서 구현될 3차원 좌표의 P_A 를 설정할 수 있다. 하지만 3개 혹은 그 이상의 수의 카메라를 통해 점을 구할 경우, P_A^1 와 P_A^2 , P_A^3 과 같이 3개, 혹은 그 이상의 점이 나오게 되므로 각 점의 가중치를 설정하는 방법이 필요하다. 본 연구에서는 3개의 카메라를 사용했을 때의 상황만을 추가적으로 고려하였다.

그림 3.5는 3개의 카메라를 통해 3차원 좌표를 구할 때의 방법을 나타낸다. 본 연구에서는 두 개의 붉은 선간의 거리를 3차원 점을 지정할 때의 가중치로 사용하였다. 식 (9)는 세 개의 점이 주어졌을 때 3차원 좌표를 구하는 나타낸다.

$$P_A = \frac{\frac{1}{|P_{i1} - P_{i2}|} P_i + \frac{1}{|P_{j1} - P_{j2}|} P_j + \frac{1}{|P_{k2} - P_{k3}|} P_k}{\frac{1}{|P_{i1} - P_{i2}|} + \frac{1}{|P_{j1} - P_{j2}|} + \frac{1}{|P_{k2} - P_{k3}|}} \quad (9)$$

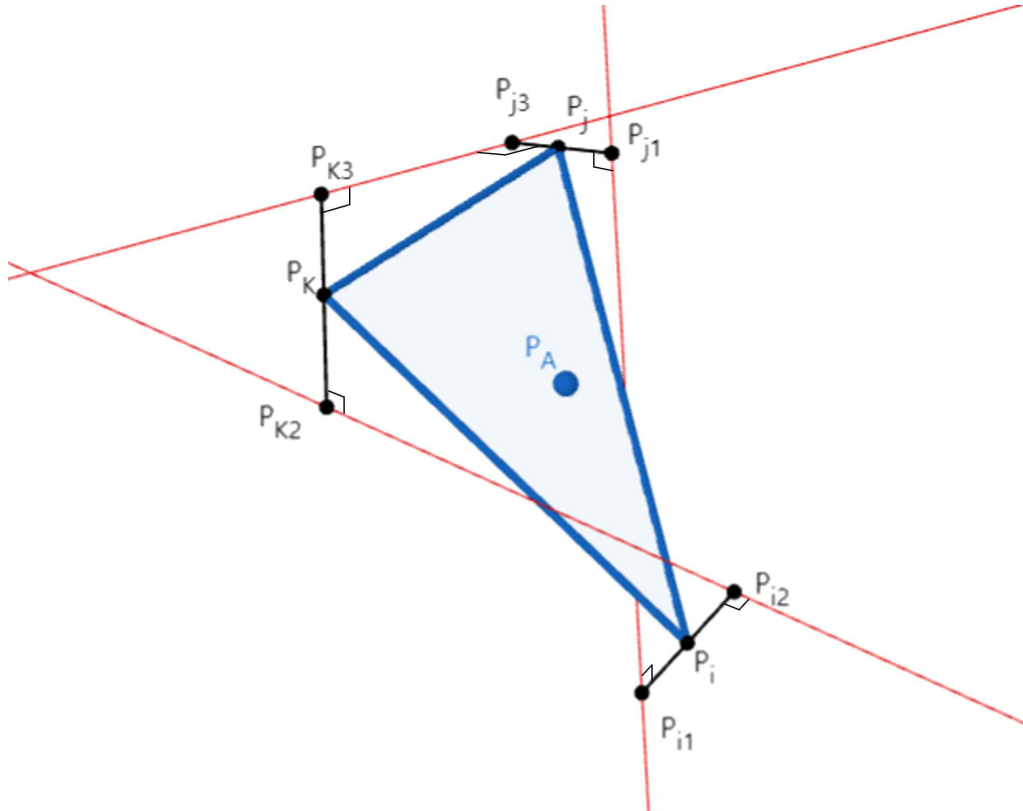


그림 3.3 3개의 카메라를 사용했을 때 3차원 좌표

이후 표현된 3차원 좌표는 사람의 움직임을 선형으로 가정하여 칼만 필터(Kalman Filter) [7]를 사용하여 보정하였다. 이후 저역 통과 필터(Low Pass Filter)를 사용하여 좀 더 부드러운 움직임이 반영되도록 설정했다. 식 (10) 해당 과정을 나타낸다. $\overline{P_A^t}$ 는 최종적으로 추정된 타임 스텝 t순간의 관절의 3차원 좌표, $\overline{P_A^t}$ 는 칼만 필터를 통해 추정된 값, α 는 이전 관절의

좌표 현재 관절의 좌표의 비중을 선택하는 파라미터이다. 값이 작을수록 현재 추정된 좌표의 비중이 커지고, 반대로 값이 커질수록 이전 위치에 대한 비중이 증가한다. 그림 3.4은 최종적으로 디지털 트윈 내에 표현된 관절의 모습을 나타낸다. 흰색 점은 관절의 위치, 붉은 선은 관절 간을 잇는 선을 나타낸다.

$$\overline{P}_A^t = \alpha \overline{P}_A^{t-1} + (1-\alpha) \overline{P}_A^{t-1} \quad (10)$$

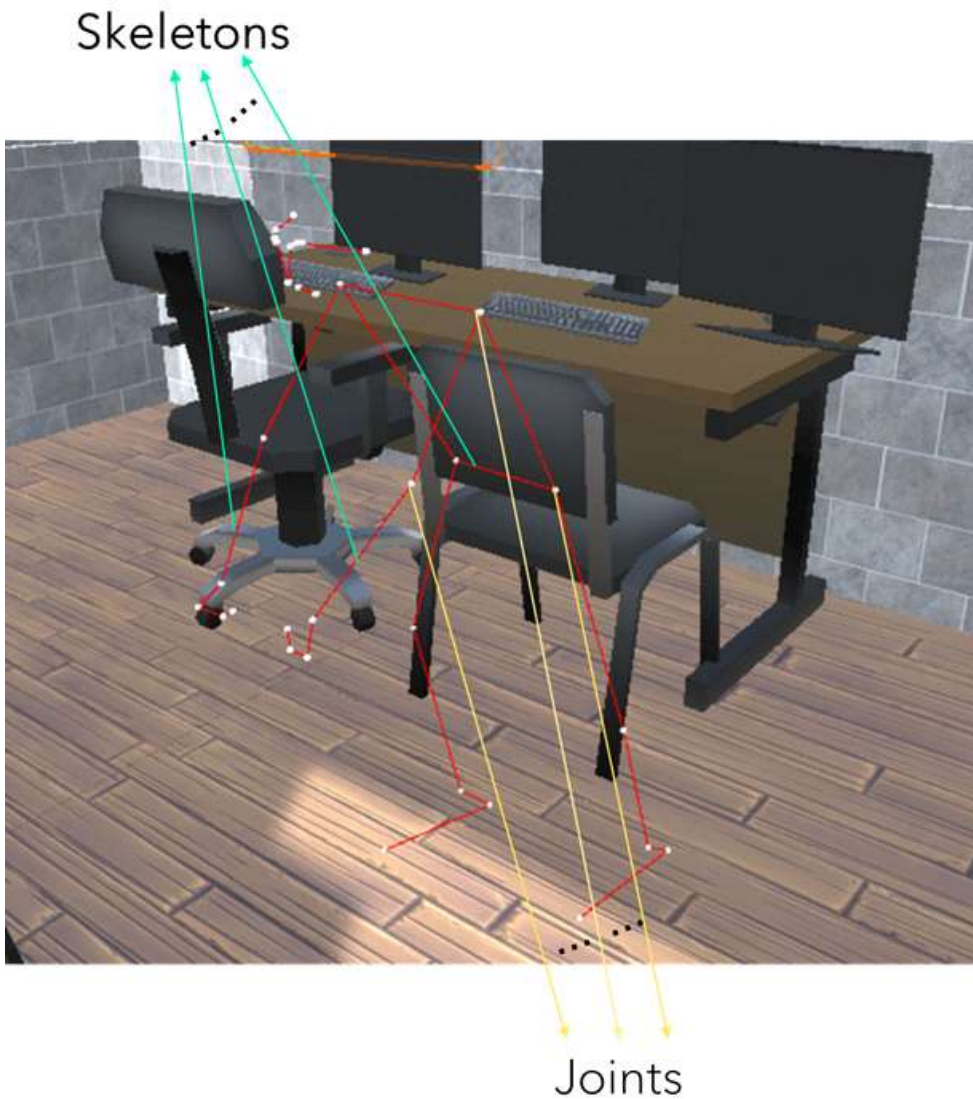
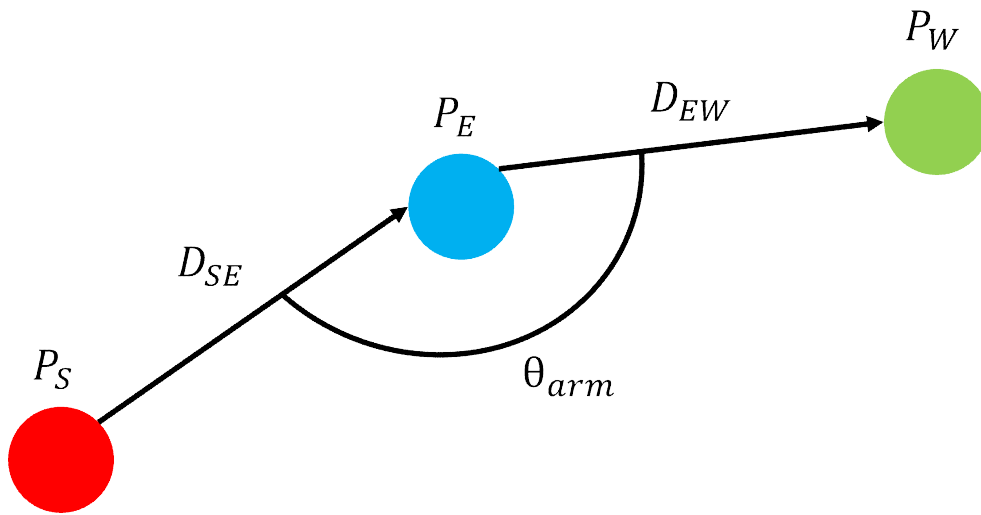


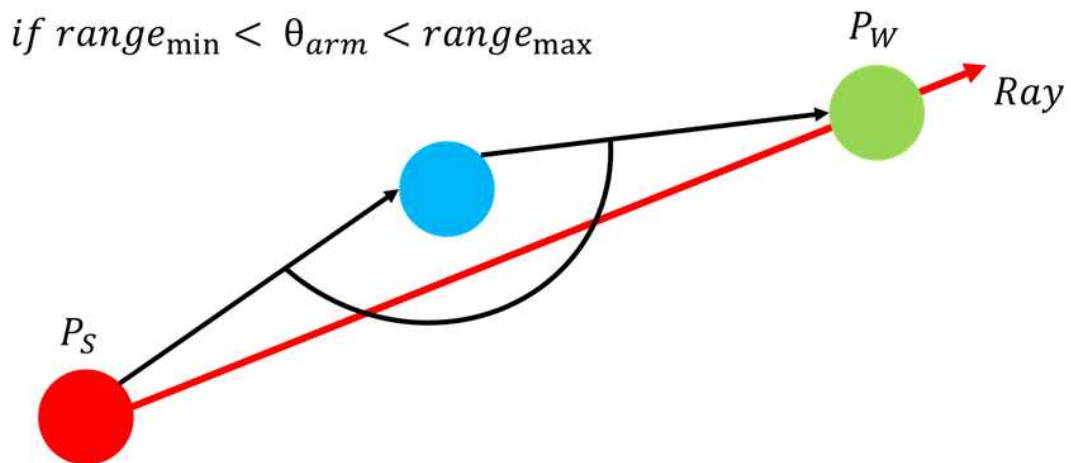
그림 3.4 3차원 좌표계에 투영된 관절

3.3.2 상호작용 구현

본 연구에서는 양 팔의 각도를 통해 상호작용의 트리거와 객체를 상호작용을 일으킬 객체로 설정했다. 3차원의 관절 위치가 디지털트윈에 표현된 상황에서 어깨의 관절의 위치를 찾는다. 어깨의 위치에서 팔꿈치의 위치를 찾아 어깨에서 팔꿈치로 이어지는 거리 벡터를 얻는다. 이후 다시 손목의 위치를 찾아 팔꿈치에서 손목으로 이어지는 거리 벡터를 계산한다. 그림 3.5은 다음의 과정을 표현한다. 어깨의 위치를 P_S , 팔꿈치의 위치를 P_E , 손목의 위치를 P_W 로 표현한다. 그러면 어깨에서 팔꿈치의 거리 벡터는 식 (11)을 통해 정의되고, 팔꿈치에서 손목 거리 벡터는 식 (12)을 통해 정의된다.



(a) 관절의 각도 계산



(b) 어깨에서 손목 방향으로 Ray를 발사
그림 3.5 상호작용 과정

$$D_{SE} = P_E - P_S \quad (11)$$

$$D_{EW} = P_W - P_E \quad (12)$$

이후 두 개의 거리 벡터를 내적을 이용한 식 (12)을 통해 다음과 D_{SE} 와 D_{EW} 사이의 각도를 계산한다.

$$\theta_{arm} = \frac{D_{SE} \cdot D_{EW}}{|D_{SE}| |D_{EW}|} \quad (13)$$

이 각도를 특정 각도 영역 안에 있는지 판단해 팔의 관계된 관절의 위치가 직선이 되는지 판단한 후 직선으로 판단되었으면 디지털 트윈 내에서 어깨에서 손목까지 무한히 이어지는 Ray를 발사한다. 해당 Ray와 HCI를 발생시킬 객체가 충돌하면 해당 객체에 사전 정의된 상호작용을 발생시킨다. 그림 3.6 실제 팔의 관절들을 통해 상호작용이 발생하는 과정을 보여준다. 어깨에서 손목으로 이어진 검은 선이 발사된 Ray를 나타내며 기존에 꺼져있던 불이 Ray와 충돌하면서 상호작용이 발생하여 점등 된 것을 확인할 수 있다.



(a) 팔이 직선으로 퍼진 것을 감지



(b) 팔의 방향으로 Ray를 발생해 상호작용 발생
그림 3.6 디지털 트윈에서의 상호작용 발생 과정

4. 실험

4.1 실험 환경

4.1.1 시뮬레이션 환경

실험은 3D 게임엔진인 Unity에서 제작한 시뮬레이션 환경에서 진행했다. 규격은 아래와 같이 실제 높이를 포함한 모든 규격을 직접 측정하여 설정했다. 총 5개의 상호작용이 가능하도록 구현했으며 2개의 문, 2개의 전등, 1개의 냉장고로 구성된다. 해당 상호작용이 수행되는 객체의 위치와 카메라의 위치는 노란색 전등, 문, 전원, 카메라 아이콘을 통해 그림 4.1과 같이 표시했다.

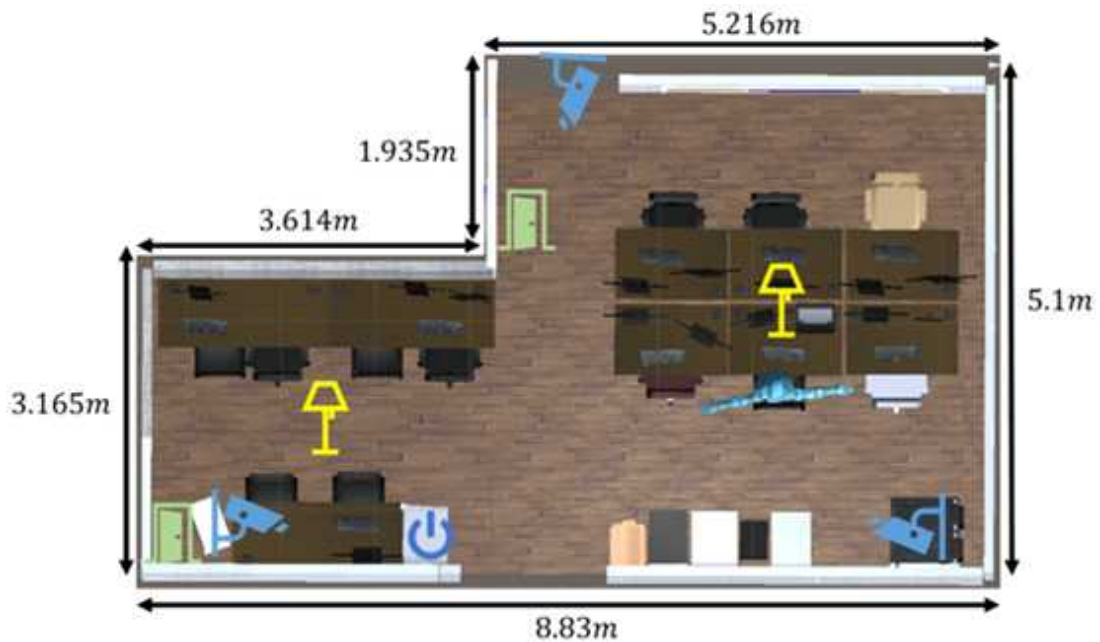


그림 4.1 Top-view 시점에서 본 실험 환경

4.1.2 상호작용 설정

상호 작용의 조건은 팔의 각도가 170도에서 180도에서 존재할 때 팔이 펴졌다고 판단하여 Ray를 어깨에서 손목으로 발사되도록 설정했다.

4.1.3 데이터 통신 및 인공지능 모델

데이터통신은 구글에서 만든 데이터베이스 플랫폼인 Firebase를 사용했다. Firebase는 실시간성이 뛰어난 데이터베이스 중 하나로 데이터의 형태는 JSON 형태를 주로 사용한다. 실제 디지털 트윈이 구현되는 PC와 자세 추정이 이루어지는 PC가 다르므로 두 PC 간의 통신을 위해서 사용했다. 3개의 2차원 관절 좌표가 Firebase에서 구축한 데이터베이스에 저장되면 디지털 트윈이 동작하는 PC에서 데이터를 종합하여 3차원 관절 좌표를 추정하는 과정을 수행했다.

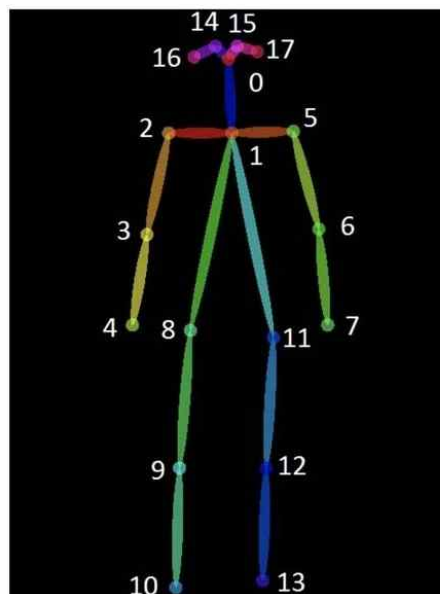


그림 4.2 Firebase

객체 검출 모델은 속도와 자세 추정 모델은 두 기능을 동시에 수행할 수 있는 YOLOv7 Pose [8] 를 사용했다. COCO 데이터셋 [9] 중 객체 검출용 데이터셋과 자세 추정용 데이터셋을 사용하여 학습된 모델을 사용했다. 추정하는 관절의 종류와 수는 표 4.1과 같으며 표 4.1의 오른쪽 그림은 각 관절의 위치와 인덱스를 나타낸다.

표 4.1 추정하는 관절의 종류

Index	Joint
0	코
1	왼쪽 눈
2	오른쪽 눈
3	왼쪽 귀
4	오른쪽 귀
5	왼쪽 어깨
6	오른쪽 어깨
7	왼쪽 팔꿈치
8	오른쪽 팔꿈치
9	왼쪽 손목
10	오른쪽 손목
11	왼쪽 골반
12	오른쪽 골반
13	왼쪽 무릎
14	오른쪽 무릎
15	왼쪽 발목
16	오른쪽 발목



4.1.4 카메라와 PC 환경

3개의 카메라는 동일 모델로 G-GOON GPRO-QHD800 4K 모델을 사용했다. 해상도는 3264 x 2448 이다. 해당 해상도를 사용하면 연산량이 커져 3개의 영상에 대해서 인공지능 모델을 사용하면 실시간성 확보가 어렵기 때문에 640 x 480의 해상도를 조정하여 사용했다.



그림 4.3 G-GOON GPRO-QHD800 4K

디지털 트윈 PC와 자세추정 PC 성능은 표 4.2에 기입한 성능으로 실험을 진행했다.

표 4.2 디지털 트윈 PC와 자세추정 PC 성능

	디지털 트윈 PC	자세추정 PC
CPU	Intel(R) Core(TM) i9-10900K CPU @ 3.70GHz	Intel(R) Core(TM) i5-12600K CPU @ 4.90GHz
OS	Windows 10	Ubuntu 20.04
GPU	NVIDIA GeForce RTX 3090	NVIDIA GeForce RTX 3080
RAM	32GB	32GB

4.1.4 카메라 파라미터

그림 4.4는 3개의 카메라 시점에서의 실제 환경과 디지털 트윈을 나타낸다. 각 카메라는 겹치는 부분이 있을 수 있도록 설정했고 사람의 신체를 전체적으로 담기 위해 높은 높이에서 아래로 내려다보는 시점으로 설치했다.



그림 4.3 3개의 카메라 시점에서 실제 환경과 디지털 트윈

표 4.3은 카메라의 내부, 외부 파라미터를 나타낸다. 세부적인 카메라 파라미터의 경우 실험적으로 초점 거리, 위치, 회전각도와 센서 사이즈를 실험적으로 조정하여 실제 카메라의 시점과 디지털 트윈에서의 카메라 시점을 유사하게 만든 후 디지털 트윈에서의 카메라 파라미터를 실제 수식에 적용했다. 카메라의 외부 파라미터는 디지털 트윈 내에서 가상의 원점을 설정하여 외부 파라미터를 위치는 미터(m) 단위, 회전 각도는 오일러 표현법으로 도(deg) 단위로 나타냈다.

표 4.3 카메라의 파라미터

카메라 ID 번호	초점 거리 (mm)	센서사이즈 (x, y) (mm)
1	9.7	(13.3, 13)
2	10.3	(13.3, 10)
3	12	(13.3, 10)

(a) 내부 파라미터

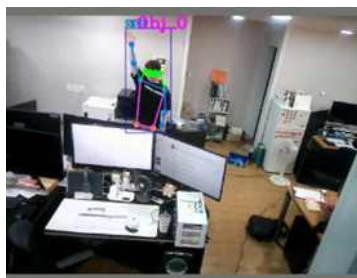
카메라 ID 번호	위치 (x, y, z) (m)	회전 각도 (x, y, z) (deg)
1	(2.309, 0.848, 3.458)	(22.415, 187.035, 3.544)
2	(4.597, 0.36, 1.393)	(8.718, -73.277, 1.393)
3	(-2.622, 0.4191, -1.1081)	(10.911, 57.232, -3.779)

(b) 외부 파라미터

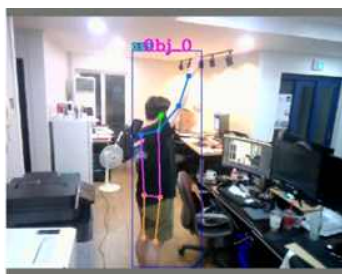
4.2 실험 결과

4.2.1 결과 사진

그림 4.4은 사람의 검출 및 자세 추정에 관한 결과 사진이다. 1번, 2번, 3번에 관한 결과 이미지도. 앞서 말한 것과 같이 검출된 사람의 바운딩박스 내부에서 사람의 관절 위치가 추정된 것을 확인할 수 있다.



(a) 1번 카메라



(b) 2번 카메라



(c) 3번 카메라

그림 4.4 자세 추정 사진

그림 4.5는 3개의 카메라에서 포착한 2차원 좌표를 통해 추정한 3차원 관절 좌표를 스켈레톤 형태로 표현한 것이다.

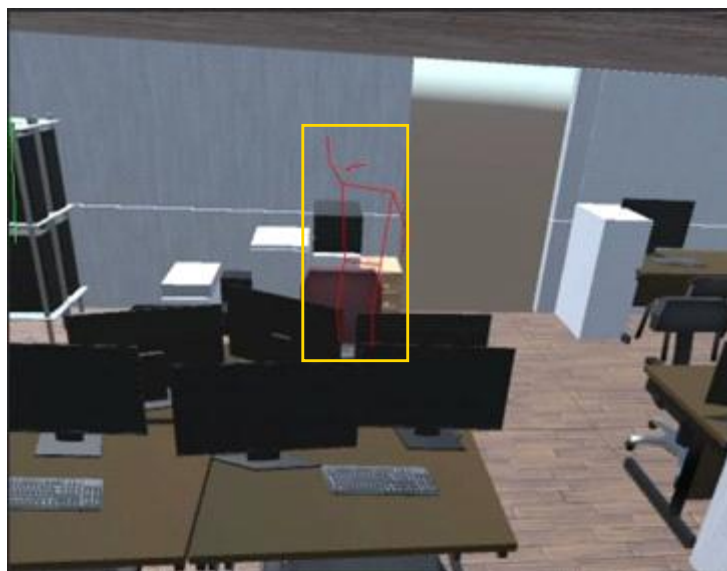


그림 4.5 투영된 3차원 관절 위치

그림 4.6은 인간형 3D 캐릭터를 사용하여 사람의 현재 상태를 표현한 것이다. 본 연구에서는 골반 관절의 중심점을 사람의 위치를 대표하는 지점으로 설정했다. 각 관절에 이어지는 방향 벡터를 개별적으로 계산해 해당 각도로 관절을 회전시키는 방법으로 캐릭터의 상태를 표현시켰다.



(a) 1번 카메라

(b) 2번 카메라

(c) 3번 카메라

그림 4.6 각 관절의 위치를 통한 캐릭터 투영

4.2.2 상호작용 성능

디지털 트윈 내 주변을 움직이며 상호작용을 15회씩 수행했으며 2개의 카메라를 사용할 때와 3개의 카메라를 사용할 때의 5개의 상호작용에 대해 성공률을 측정하고 비교했다. 상호작용 시도 후 3초 이내에 상호작용이 성공하면 대해서만 성공으로 간주했다. 표 4.3은 카메라 2개를 사용했을 때 성능을 나타낸다.

표 4.4 카메라 2개를 사용했을 때 상호작용 성능 (%)

1번 문	2번 문	1번 전등	2번 전등	냉장고
46.67	69	66.67	46.67	86.67

두 개의 카메라만 사용한 경우, 일정 객체에 대해서 일부 관절이 가려지는 것이 원인이 되어 1번 문과 2번 전등에 대해서 상호작용을 시도할 때 낮은 성능을 보였다

표 4.5 카메라 3개를 사용했을 때 성능 (%)

1번 문	2번 문	1번 전등	2번 전등	냉장고
80	66.67	86.67	80	100

표 4.5는 카메라 3개를 사용했을 때의 상호작용 성능을 나타낸다. 1번 문과 2번 전등에 대해서 2개의 카메라를 사용했을 때보다 모두 33.33% 상승한 것을 보여준다. 이외에도 모든 상호작용 객체에 대해서 평균적으로 21.33%의 상승이 이루어진 것을 확인할 수 있다.

5. 결론

스마트 홈은 현재 국내, 국외 시장 모두에서 주목받고 있다. 스마트 홈을 구축하기 위해서 우선시 되는 것은 HCI를 구축하는 것으로, 구축하는 방법에 대해서는 다양한 접근 방식이 있다. 본 논문은 각 방법에 대한 문제를 제기하고, 문제를 해결하기 위한 HCI 구현 방법을 설명한다. 본 연구에서는 실내 건물에 설치된 복수의 카메라로 촬영된 영상을 통해 사람의 자세를 실시간으로 추정하고, 추정된 2차원 관절의 위치를 통해 3차원 관절 위치를 설정해 팔의 방향과 퍼짐 정도를 인지하여 착용 및 물리적인 별도의 장비 없이 HCI 시스템을 구현 및 개발하는 연구를 수행했다. 단계는 디지털 트윈 구축, 사람의 검출 및 자세 추정, 3차원 좌표 설정, 상호작용 발생으로 이루어진다. 실험에서는 두 개의 카메라와 세 개의 카메라를 사용할 때의 상호작용 성능에 대해 평가했고 각각 63.6%, 82.7%의 성능을 거두었다. 향후, 객체 검출과 자세 추정 모델에 대한 성능을 증진하면 상호작용에 대한 성능이 향상할 것으로 전망된다. 또한, 컴퓨터비전 분야 중 하나인 Re Identification을 사용하여 특정 사용자만 인식하는 기능을 추가한다면 조금 더 사용자에게 친화적인 HCI 시스템을 구현할 수 있을 것으로 생각된다. 또한, 데이터베이스 플랫폼을 이용하는 것이 아닌 다른 프로토콜을 사용해서 데이터 통신이 이루어진다면, 좀 더 빠른 속도로 3차원 자세 추정이 가능할 것으로 생각된다.

참고문헌

- [1] Hartley, Richard, and Andrew Zisserman. Multiple view geometry in computer vision. Cambridge university press, 2003.
- [2] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
- [3] Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." Advances in neural information processing systems 28 (2015).
- [4] Cao, Zhe, et al. "Realtime multi-person 2d pose estimation using part affinity fields." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
- [5] Thelen, Adam, et al. "A comprehensive review of digital twin—part 1: modeling and twinning enabling technologies." Structural and Multidisciplinary Optimization 65.12 (2022): 354.
- [6] Grieves, Michael W. "Product lifecycle management: the new paradigm for enterprises." International Journal of Product Development 2.1-2 (2005): 71-84.
- [7] Kalman, R. (1960) A New Approach to Linear Filtering and Prediction Problems. ASME Journal of Basic Engineering, 82, 35-45.
- [8] Wang, Chien-Yao, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023. - YOLOv7
- [9] Lin, Tsung-Yi, et al. "Microsoft coco: Common objects in context." Computer Vision-ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. Springer International Publishing, 2014.