

# AI 비전과 생성형 AI 를 이용한 멀티 홈 디바이스 제어

홍수민, 김수민, 송수희, 안채연  
이화여자대학교 전자전기공학전공 학부생

hsum916@ewhain.net, als000620@ewhain.net, songsoo0037@ewhain.net, jies0755@ewhain.net

## Control of Multi-Home Devices Using AI Vision and Generative AI

Su-Min Hong, Su-Min Kim, Su-Hee Song, Chae-Yeon Ahn  
Dept. of Electronic and Electrical Engineering, Ewha Womans University

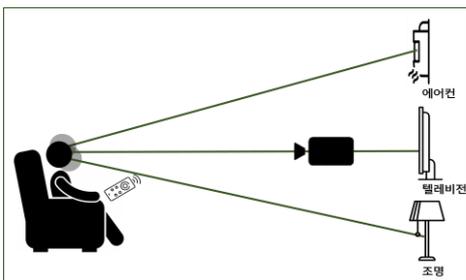
### 요 약

기술의 발전으로 인해 스마트 가전제품이 늘어나며 스마트 홈 기술이 주목을 받고 있다. 그러나 이러한 기술은 설정과정의 복잡성으로 사용자들이 쉽게 접근하기 어렵다. 특히 디지털 기기 사용에 익숙하지 않은 사용자들을 스마트 홈 기술로부터 소외시키는 결과를 낳고 있다. 본 논문에서는 사용자 친화적인 스마트 홈 시스템을 제안한다. 사용자의 시선 방향을 추적하여 디바이스를 선택하고 간단한 인터페이스의 컨트롤러로 디바이스를 손쉽게 조작할 수 있도록 한다. 또한, 생성형 인공지능과 RAG 를 결합하여 사용자가 가전제품과 자연스럽게 대화하며 정보를 얻을 수 있는 인터페이스를 제공한다.

### 1. 연구 동기 및 배경

가정 내 전자기기의 수는 기술의 발전으로 인해 점차 증가하고 있으며, 이러한 기기들의 다양성도 함께 확장되고 있다. 이전에는 일반적인 가전제품이 주를 이루었지만, 현재에는 사물 인터넷(IoT) 기술을 통합한 스마트 가전제품들이 등장하고 있다. 이로 인해 가전 기기를 제어하는 방법 또한 다양화되어 왔으며, 이러한 기기들을 연결하고 제어하는 주요 방법의 하나로 IoT 플랫폼을 기반으로 하는 스마트 홈 애플리케이션이 널리 쓰이고 있다.

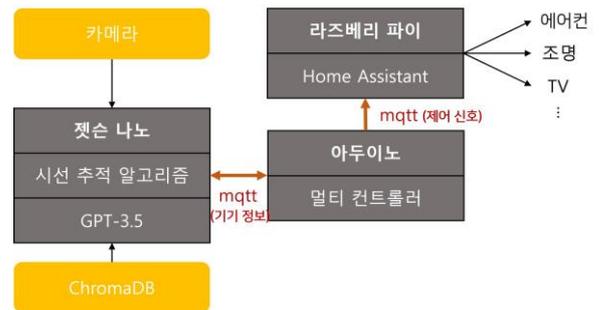
그러나 모든 사용자가 이러한 기술의 발전에 쉽게 접근 가능한 것은 아니다. 정보 격차는 주로 IoT 서비스 설정 과정의 복잡성으로 인해 발생한다. 스마트 홈 서비스 설정 과정이 사용자에게 특정한 기술 지식과 경험을 요구하기 때문에 디지털 기기 사용에 익숙하지 않은 디지털 소외계층은 스마트 홈 기술의 혜택을 충분히 누리기 어려운 상황에 처해 있다.[1]



(그림 1) 스마트 홈 시스템 사용 환경

이에 본 논문에서는 사용자의 시선을 추적하여 제어할 기기를 선택해주는 사용자 친화적 스마트 홈 시스템을 제안한다.

### 2. AI 를 이용한 멀티 홈 디바이스 제어



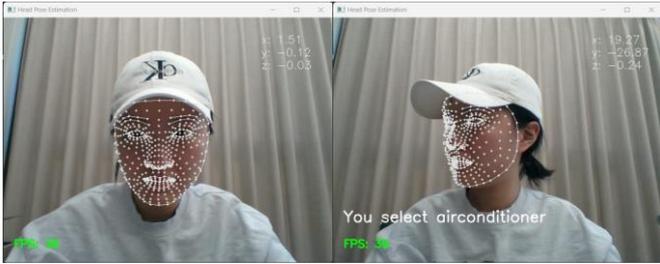
(그림 2) AI 를 이용한 스마트 홈 시스템 구조도

#### 2.1.A. 시선 방향 추적을 통한 기기 선택 알고리즘

카메라와 사용자 사이의 거리가 멀거나 사용자의 눈이 다른 물체로 가려져 있는 경우, 눈 정보를 사용하여 시선을 추적할 수 없다. 사람의 시선 방향과 얼굴 방향 사이에는 깊은 연관성이 있기 때문에 얼굴 방향 벡터를 계산해 시선 방향을 추정하기로 결정했다.[2]

사용자가 제어하고 싶은 전자기기를 바라보며 컨트롤러를 조작하면 웹캠과 연결된 젯슨 나노(Jetson Nano) 보드에서 사용자의 얼굴 이미지를 캡처한다.

시선 방향 추적 알고리즘은 Google AI 프레임워크 Mediapipe 의 face mesh 라이브러리를 이용해 얼굴 특징점을 추출하고, PnP 알고리즘을 이용해 이미지 좌표계에서 실제 좌표계로 변환한다. 그 후, 얼굴 방향 벡터를 계산해 사용자가 현재 어떤 디바이스를 바라보고 있는지 알아낸다. 선택된 디바이스 정보는 Publish-Subscribe 기반의 메시지 송수신 프로토콜인 MQTT(Message Queuing Telemetry Transport)를 통해 컨트롤러로 전송되며, 사용자는 컨트롤러를 이용해 해당 디바이스를 바로 제어할 수 있다.



(그림 3) 시선 방향 추적을 이용한 기기 선택 과정

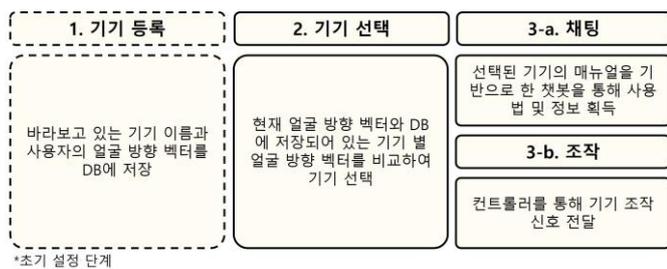
2.1.B. 시선 방향 추적 알고리즘 정확도 테스트

벽면을 9 개 구역(3x3)과 15 개 구역(3x5)으로 나누어 두 번의 실험을 진행했다. 각 구역을 처음에 바라본 시선 각도를 정답으로 설정하고, 각 실험마다 100 번씩 시선 각도 테스트를 진행하였다. 9 개 구역 실험에서는 90.2%의 정확도와 9.091 의 RMSE 값을 얻었으며, 15 개 구역 실험에서는 94.1%의 정확도와 8.420 의 RMSE 값을 얻었다.

위 실험을 통해 해당 알고리즘이 가전기기를 선택하기에 충분히 높은 정확도로 시선 방향을 구분한다는 것을 확인할 수 있었다.

2.2. RAG 를 접목한 LLM 모델

생성형 AI 와 RAG 를 활용한 챗봇 시스템을 통해 사용자는 가전제품과 대화하며 사용법 및 기기에 대한정보를 알아낼 수 있다. RAG(Retrieval Augmented Generation)란 외부 정보를 미리 저장한 후 질의가 들어오면 관련된 정보를 검색하여 해당 정보가 담긴 문서들을 프롬프트를 통해 LLM 에 전달하는 방식이다. GPT-3.5 를 이용해 가전제품들의 매뉴얼을 벡터 형태로 변환한 후, 이를 Chroma DB 에 저장하였다. 사용자가 한 질문을 임베딩 벡터로 만들고 벡터 DB 에서 이와 가장 유사도가 높은 벡터를 찾아 답변을 제공하기 때문에 사용하는 기기에 가장 적합한 정보를 얻을 수 있다.



(그림 4) 스마트 홈 시스템 사용 시나리오

3. 멀티 홈 디바이스의 매니지먼트 시스템

3.1. 멀티 디바이스 컨트롤러

Arduino Nano 33 IoT 와 모듈을 연동하여 멀티 디바이스 컨트롤러를 구현했다. 초기 단계에서 회로 연결의 신뢰성을 평가하기 위해 bread board 를 활용해 프로토타입을 제작했다. Bread board 를 활용한 회로 개발은 사용하는 모듈의 개수가 늘어남에 따라 신호가 불안정해지는 문제가 있다. 전체 시스템의 안정성을 최적화하고 신뢰성을 보장하기 위해 오픈소스 CAD 프로그램 Fritzing 을 이용해 PCB 회로를 설계 및 제작하고 구성 모듈을 납땜해 집적화 프로세스를 진행하였다.

3.2. Home Assistant 서버

라즈베리 파이(Raspberry Pi)에 Home Assistant(HA) OS 를 설치해 스마트 홈 서버를 구축했다. TV 와 에어컨은 와이파이(WI-FI) 연결, 조명은 지그비(ZigBee) 연결을 통해 통합된다. HA 는 특정 trigger 가 발생하면 미리 설정된 action 을 수행하는 자동화 설정을 할 수 있다. 이를 이용해 컨트롤러에서 MQTT 를 통해 HA 서버로 기기 정보와 제어 신호가 전달되면 자동화 기능을 통해 기기를 조작할 수 있게 했다.

4. 결론

본 연구는 스마트 홈 기술의 접근성을 높이기 위한 두 가지 주요한 접근 방식을 제시하고 있다.

첫째로, 시선 방향 추적 알고리즘을 통해 사용자가 바라보는 디바이스를 알아내어 기존의 복잡한 어플리케이션의 설정 과정 없이 손쉽게 가전제품을 제어할 수 있다. 직접 실험을 수행하여 해당 알고리즘이 높은 정확도를 보임을 확인하였다.

둘째로, 생성형 AI 와 RAG 를 접목한 LLM 모델을 활용하여 사용자가 마치 가전제품과 대화하듯 소통하며 사용법에 대한 정보를 얻을 수 있도록 하였다. 길고 복잡하여 읽기 어려운 기존의 매뉴얼에 담겨있는 정보를 시각이나 문해 능력에 제한이 있는 사용자 또한 쉽게 활용할 수 있을 것이다.

이 두 가지 접근 방식은 디지털 소외계층을 포함한 모든 사용자가 스마트 홈 기술을 효과적으로 활용할 수 있는 기회를 제공할 것이다. 또한, 멀티 컨트롤 매커니즘은 가정 이외에도 학교나 병원 등 다양한 기기를 편하게 제어할 필요가 있는 장소에서 유용하게 사용될 것으로 기대된다.

본 논문은 과학기술정보통신부 정보통신창의인재양성사업의 지원을 통해 수행한 ICT 멘토링 프로젝트 결과물입니다.

참고문헌

[1] K Rasheva-Yordanova, "IOT: A NEW DIMENSION OF DIGITAL DIVIDE", ICERI2021 Proceedings, 2021, pp. 1811-1816.

[2] J Leroy, "Second screen interaction: An approach to infer TV watcher's interest using 3D head pose estimation", Proceedings of the 22nd International Conference on World Wide Web, New York, NY, USA, 2013, pp.465-468.