

현실 세계를 제어하는 혼합 현실 대화 시스템

김다혜⁰, 박규훤, 정영섭
순천향대학교 빅데이터공학과

ekgp908@gmail.com, {svclaw2000,bytecell}@sch.ac.kr

Mixed Reality Dialog Agent for Real-World Control

DaHye Kim⁰, KyuHwon Park, Young-Seob Jeong
Big Data Engineering Department, Soonchunhyang University

요 약

사용자와 자연어 음성을 통해 상호작용하는 대화 시스템이 주목받고 있다. 음성 뿐 아니라, 화면을 통해 사용자와 상호작용하는 기능을 제공하는 대화 시스템들이 소개되고 있으며, 최근에는 혼합 현실 환경에서 대화 에이전트가 실제 공간에 존재하는 서비스가 소개되기도 하였다. 본 연구에서는 혼합 현실 환경의 대화 시스템이 현실 세계에 영향을 미칠 수 있는 시스템을 설계하였다. 제안하는 시스템은 서버가 클라이언트와 통신하여 필요 시 적절한 장치를 제어하도록 설계되었다. 본 연구 시스템의 실현가능성을 입증하기 위해 라즈베리파이를 제어할 장치로 사용하였으며, 화면에 보이는 대화 에이전트에게 음성으로 명령하여 전등, 에어컨 등을 제어하는 것이 가능함을 확인하였다.

주제어: 대화시스템, 챗봇, 혼합 현실, 현실 세계 제어

1. 서론

최근, 사람과 기계가 자연어를 통해 상호작용하는 대화 시스템이 빠르게 성장하고 있다. 대부분의 대화 시스템은 음성 또는 텍스트를 통해 상호작용 할 수 있지만, 영상을 통해 상호작용하는 기능을 제공하는 대화시스템은 비교적 적다. 최근에는 혼합현실(Mixed Reality) 환경에서 사용자와 상호작용하는 시스템이 개발되고 있는데, 본 연구는 혼합 현실 환경에서 대화 에이전트에게 명령하여 현실 세계에 영향을 줄 수 있는 시스템을 설계하였다.

본 연구의 대화 시스템은 서버가 클라이언트와 장치(gadget)들을 관리하는 역할을 수행하도록 설계되었다. 클라이언트는 다양한 센서로부터 실시간으로 데이터를 수집하고, 이 데이터를 기반으로 사용자의 방향을 계산하고 실내 위치를 서버로부터 전달받는다. 서버는 클라이언트로부터 사용자 입력(음성 또는 텍스트)을 전달받고, 이를 자연어 처리 기술을 활용하여 적절한 시스템 응답을 생성한다. 특히, 클라이언트의 현재 실내 위치를 계산하기 위해 k-Nearest Neighbors(kNN) 알고리즘을 사용하였다. 본 연구의 실험에서는 라즈베리파이를 장치로 사용하였으며 서버와 통신을 하기위해 CherryPy 라이브러리를 사용했다.

본 연구의 시스템을 기존의 스마트홈, 대화시스템들과 비교했을 때 가지는 차별성은 다음과 같다. 첫째, 장치들을 원하는 곳에 설치함으로써 다양한 동작 수행하도록 손쉽게 확장이 가능하다. 둘째, 대화 에이전트가 실내 공간에 실재하므로 사용자는 클라이언트 화면을 통해 실내 위치에 존재하는 에이전트를 볼 수 있으며, 에이전트에 충분히 접근하여 대화하거나 명령을 내릴 수 있다. 이 때, 현실 세계에 영향을 미치는 명령, 이를테면 전등 스위치를 켜거나 끄는 등의 명령을 내릴 수 있으며, 현

실 세계에 보이는 대화 에이전트가 현실 세계에 영향을 미치는 것과 같은 경험을 제공하게 된다.

본 논문은 다음과 같이 구성되어 있다. 2장에서는 연구 배경, 필요성을 설명하고, 3장에서는 본 연구의 대화 시스템 구조에 대해 설명한다. 4장에서는 본 연구에서 제안하는 대화 시스템의 설치, 동작 결과에 대해 설명하고, 5장에서는 결론을 맺고 향후 연구 계획을 설명한다.

2. 연구 배경

최근 국내에서는 4차 산업혁명 ICT 기술인 가상현실(Virtual Reality)과 증강현실(Augmented Reality)에 대한 관심이 높아지고 있다. 이에 따라, IT업계는 VR/AR 시장을 차세대 컴퓨팅 플랫폼으로 전망하고 있으며, 게임, 엔터테인먼트뿐만 아니라 마케팅, 교육, 오락, 의료 등 범위가 확장되면서 다양한 콘텐츠로 널리 제작되고 있다. 가상현실은 작은 화면을 통해 현실 세계와 단절된 콘텐츠이기 때문에 몰입도가 높지만, 현실감이 부족하다. 반면 증강현실은 현실을 배경으로 하여 현실감이 있지만, 몰입도가 낮다. 다음과 같은 부분들을 개선해, 현실과 가상을 동시에 제공해주는 혼합현실(Mixed Reality)이 등장했다. 이러한 MR 환경에서 사람과 상호작용하는 시스템 또한 개발되고 있다.

최근 상호작용 기술로서 조명받는 것은 대화 시스템(dialogue system)이다. 대화 시스템은 사람과 기계가 자연어를 사용하여 상호작용을 함으로써 사용자가 원하는 목적을 성취하도록 하는 시스템을 말한다 [1]. 아마존 알렉사 [2], 구글 홈 [3], 네이버 클로바 [4] 등의 개발로 대화형 인공지능이 많은 관심을 받으며 빠르게 성장하고 있다. 하지만, 이러한 대화 시스템 대부분은 음성 또는 텍스트를 통해 상호작용할 수 있으며, 그림, 영상을 통한 상호작용은 제공되지 않는다는 한계가 있

다. 이러한 대화형 인공지능의 단점을 극복한 다양한 연구들이 등장했다.

최근, 대화 시스템을 통한 상호작용 서비스를 혼합현실 환경에서 제공하는 사례가 등장하고 있다. 예를 들어, PhoenixBot은 실제 물리적 공간에서 위치에 있는 대화형 에이전트가 가까이 위치한 사람과 상호작용하게 해준다 [5]. PhoenixBot의 에이전트 위치는 마커에 의해 결정되지 않아 사용자와 에이전트 사이의 거리가 충분히 가까울 때 통신한다. 즉, 시선이 가까운 물체에서 먼 물체로 옮겨질 때 초점도 같이 이동하므로 생생한 체험이 가능하다. 하지만, PhoenixBot의 대화형 에이전트가 현실 세계에 간섭할 수 없다는 한계가 있었다.

본 연구에서는 [5]의 대화 시스템을 확장하여, 혼합현실 환경에서 현실 세계에 영향을 미칠 수 있는 대화 시스템을 설계하였다. 가령, 대화형 에이전트에게 명령하여 연구실 내 전등을 켜고 끄거나, 가정집에서는 출입문을 여닫는 데 사용할 수 있을 것이다.

현재 상용화된 스마트 홈 시스템도 본 연구와 비슷한 기능을 제공한다. 스마트 홈 시스템과 본 연구의 대화 시스템의 가장 큰 차이점은, 스마트 홈 시스템은 최초 설계 시점부터 스마트 홈 환경을 제공하도록 설계되었을 때 대체로 가능하지만, 본 연구의 대화시스템은 기존의 주거 건물에서도 필요 장치를 설치하기만 하면 사용할 수 있다. 그뿐만 아니라, 대화 에이전트가 현실 세계의 물리적 공간에 실재하므로, 화면 등을 통해 대화 에이전트를 실내에서 볼 수 있으며, 대화 에이전트에게 명령하거나 대화를 나눌 수 있다.

3. 시스템 설계

3.1 시스템구성

본 연구에서 소개하는 시스템은 그림 1과 같이 동작한다. Client는 User가 다루는 스마트폰 등이 해당되며, Server는 Client와 Gadget들과 통신하면서 각종 사용자 입력을 처리하고 Client와 Gadget이 적절한 동작을 수행하도록 제어한다. 가령, User가 스마트폰 애플리케이션을 통해 명령을 내리면 이 명령은 Server에 전달된다. Server는 전달받은 명령을 자연어 처리 등의 기술을 통해 처리하고, 그 결과물을 Client 및 적절한 Gadget에게 보낸다. Client가 Server로부터 결과물을 받으면 여러 가지 수단(예: 스마트폰 화면)을 통해 사용자에게 결과를 전달한다. Gadget 또한 마찬가지로 전달받은 결과물에 해당하는 동작을 수행한다.

User가 어떤 방의 전등 스위치를 켜기 위해 대화 시스템을 사용한다고 가정해보자. User는 Client에게 “불켜”라는 음성 명령을 하고, 이 데이터는 Server에 전달된다. 이때, Client가 직접 음성을 텍스트로 변환한 후 텍스트 데이터를 Server에 전달하거나, 음성 데이터를 Server에 전달할 수도 있는데, 본 연구에서는 Client에서 직접 음성을 텍스트로 변환한 후, 텍스트를 Server에 전달하도록 하였다. Server에서는 자연어 처리 기술을 통해 {Gadget 1, move, arm2, +T00:00:01, Locate}와 같

이 Predicate와 Object를 담고 있는 json 형식의 결과물을 생성하고, 이를 해당 Subject인 Gadget 1과 Client에 전달한다. 해당 Gadget은 이 결과물을 해석하여 arm2(두 번째 기계 팔)를 움직이게 되고, 이 두 번째 기계 팔의 동작이 전등 스위치를 켜게 된다. 결과적으로, User는 Client의 화면 등을 통해 보이는 대화 에이전트가 현실 세계에 존재하는 전등 스위치를 켜는 경험을 하게 된다.

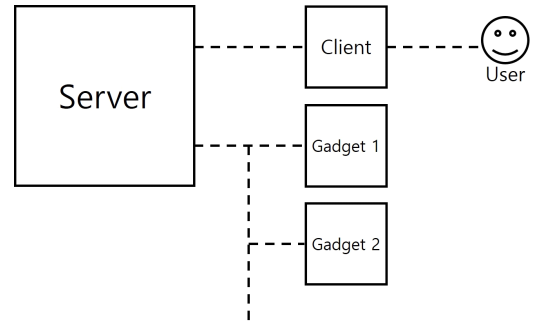


그림 1. 시스템 구조

3.2 클라이언트

본 연구에서는 클라이언트로 안드로이드 스마트폰을 사용하였다. 스마트폰의 gyroscope, accelerometer, step detector, WiFi, magnetic field, gravity sensor 등 다양한 센서로부터 실시간으로 데이터를 수집하는 애플리케이션을 개발하였다. 사용자가 스마트폰을 똑바로 들고 화면을 보고 이동한다고 가정하였으므로, 스마트폰과 사용자의 위치가 같다고 볼 수 있으며, 스마트폰의 뒷면이 향하는 방향이 사용자가 바라보는 방향이라고 할 수 있다. 스마트폰의 step detector와 WiFi 신호는 사용자의 현재 위치를 추정하는 데 사용했으며, 나머지 센서들은 사용자의 방향을 계산하는 데 사용하였다.

본 연구의 대화 시스템에서는 사용자뿐 아니라 대화형 에이전트(가상 객체)도 현실 세계에서 실제 위치를 가지고 있다. 다시 말해서, 현실 세계의 공간과 일치하는 가상 세계 맵이 존재하며, 이 맵에 사용자와 에이전트의 위치를 표현한다. 따라서, 사용자가 대화 에이전트가 위치한 곳을 바라볼 때는 스마트폰 화면에 에이전트가 보이게 되는데, 대화 에이전트가 화면에 표시될 올바른 위치를 계산하기 위해 에이전트의 정보를 서버와 일관되게 유지한다. 즉, 사용자가 이동하면 수정된 위치정보를 서버가 저장하게 되므로, 결국 모든 사용자와 에이전트의 위치정보, 방향 정보가 서버에 유지된다.

정리하자면, 클라이언트에서 스캔한 WiFi 신호를 통해 서버로부터 사용자의 현재 위치를 알아내고, 다른 센서 데이터를 바탕으로 사용자의 방향 정보를 계산하여 서버에 전송한다. 더불어, 대화 에이전트 등의 가상 객체들을 화면의 적절한 곳에 표시한다. 특히, 사용자가 대화 에이전트와 충분히 가까이 위치해 있고 에이전트를 바라보고 있을 때, 에이전트에게 명령하거나 대화할 수 있으며, 사용자 요청에 따라 현실 세계에 영향을 미치는 동작(예: 전등 스위치 켜기)을 수행한 경우에는 대화 에

이전트가 임의의 애니메이션 동작을 보이도록 하였다.

3.3 서버

클라이언트로부터 텍스트 형태의 사용자 입력을 전달 받은 서버는 자연어 처리 기술을 통해 사용자 의도, 슬롯(slot) 등을 파악한다. 이로부터 사용자에게 전달할 결과물을 생성하는데, 이는 텍스트, 음성, 그림 등의 다양한 형태로 제공될 수 있다. 본 연구에서는 안드로이드 스마트폰을 클라이언트로 가정하고 있으며, 사용자에게 전달되는 결과물은 화면에 보일 텍스트, 캐릭터 애니메이션, 음성 형태로 한정하였다.

사용자 입력으로부터 결과물을 만들어내는 역할 외에도, 서버는 해당 클라이언트의 현재 위치를 예측하는 역할도 수행한다. 클라이언트에서 전달받은 WiFi 스캔 결과를 사용하여, 이미 학습된 모델을 통해 클라이언트의 현재 실내 위치를 예측한 결과를 클라이언트에 제공한다. 본 연구에서는 이 모델에 k-Nearest Neighbors (kNN) 알고리즘을 사용하였다. 서버가 수행하는 또 다른 역할은 필요할 때 적절한 장치(Gadget)에 사용자가 원하는 동작을 수행하도록 신호를 보내는 것이다.

서버가 수행하는 역할을 한 가지 예시를 통해 보자면, 사용자 클라이언트로부터 전달받은 WiFi 신호에 기반하여 사용자 위치를 클라이언트에게 전달한다. 만약, 사용자가 대화 에이전트와 충분히 가까운 위치에 있고 사용자로부터 ‘불 켜’ 라는 요청이 전달된 경우, 이 요청으로부터 사용자 의도(예: TurnonLight) 등을 파악하여 ‘불을 켭니다’ 와 같은 결과문장을 생성한다. 특히, 전등 스위치를 켜기 위해 제어할 필요가 있는 장치를 알아내어, 해당 장치가 특정 동작(예: 로봇 팔 움직이기 등)을 수행하도록 제어함으로써 사용자가 원하는 결과인 전등 스위치를 켜는 결과를 만들어낸다.

3.4 장치

본 연구에서는 대화 에이전트가 다루는 장치(gadget)로써 라즈베리파이를 사용하였다. 라즈베리 파이는 영국의 라즈베리 파이 재단이 학교에서 기초 컴퓨터 과학 교육을 증진하기 위한 단일 칩을 사용한 저가형 싱글 보드 컴퓨터이다 [6]. 하드웨어 제어를 위한 입출력 핀인 GPIO를 내장하고 있으므로 추가적인 장비 없이도 기본적인 하드웨어 제어가 가능하다. 크기는 작지만 일반 컴퓨터처럼 다양한 기능을 수행할 수 있으며 활용 범위 또한 다양하다. 프로그래밍 언어에 대한 학습으로 사용할 수 있으며, 다양한 입출력 장치를 연결하여 활용할 수도 있다 [7].

대화 에이전트 서버가 라즈베리파이 장치를 제어하는 통신을 가능하게 하기 위해, CherryPy 라이브러리를 사용하여 라즈베리파이 장치가 신호대기(listen)하도록 하였다. 전등 스위치를 켜거나 에어컨 전원을 끄는 등의 동작을 수행하기 위해 서버로부터 두 대를 11번 핀(GPIO 17)과 12번 핀(GPIO 18)에 연결하였다.

서버로부터 전달받는 신호는 json 포맷을 따르고 있으

며 {Subject, Predicate, Object, Time, Locate} 와 같은 값들을 가진다. Subject는 동작할 장치(예: Gadget 1, Gadget2), Predicate는 동작(예: move), Object는 대상(예: arm1, arm2), Time은 시간(예: +T00:00:01), Locate는 위치를 뜻한다. Time을 나타내는 값은 ISO8601 표준을 근거하였으며, 대화 에이전트의 캐릭터 애니메이션 동작과 라즈베리파이 장치 동작의 시간차를 없애기 위해 주로 사용하였다 [8].

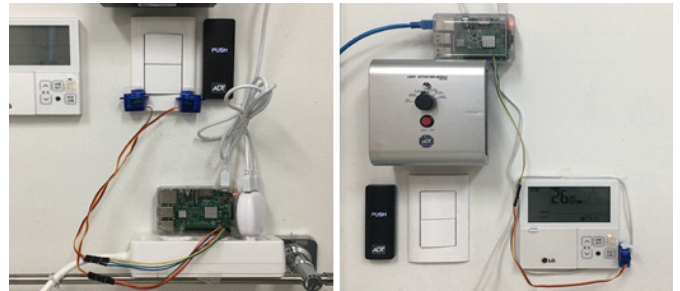


그림 2. (좌) 연구실1 전등 제어를 위해 설치한 장치, (우) 연구실2 에어컨 제어를 위해 설치한 장치

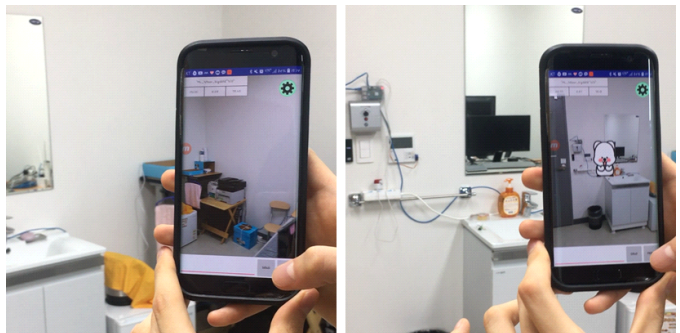


그림 3. 사용자의 방향에 따른 대화 에이전트 표시. (좌) 사용자가 에이전트를 바라보고 있지 않은 경우, (우) 사용자가 에이전트를 바라보고 있는 경우.

4. 실험

본 연구의 실험을 위해 기존 연구 [5]에서 설계했던 대화 시스템에 기능을 확장하는 방식으로 대화 에이전트를 구현하였다. 대화 에이전트는 날씨 정보 등을 제공할 수 있으며, 다양한 장치들을 제어할 수 있도록 설계되었다. 대화 시스템에서 제공 가능한 서비스 도메인은 약 7개(날씨, 지하철, 성경, 대학입시, 장치제어, 응급상황, 일반대화)이며, 본 연구의 실험은 장치제어 도메인에 해당하는 사용자 입력이 주어졌다고 가정한다.

최초 설계 시점부터 스마트 홈 환경으로 설계해야 하는 스마트 홈 시스템과는 달리, 본 연구의 대화 시스템은 필요 장치를 필요한 곳에 설치함으로써 다양한 기능을 제공 가능하다. 실험을 위해 그림2와 같이 연구실 1(305호), 연구실2(312호)에 각각 장치들을 설치했다. 연구실1에는 전등을 켜고 끄기 위해 스위치 양쪽에 장치를 설치했고, 연구실2에는 에어컨 작동을 위해 에어컨 제어기 오른쪽에 설치했다. 본 연구에서는 전등과 에어컨 제어를 위해 장치를 설치했으나, 공기청정기 제어,

컴퓨터 전원 제어, 커피 내리기 등의 다양한 동작에 응용이 가능하다.

본 연구의 대화 에이전트는 단순히 스마트폰 화면 중앙에 위치하는 것이 아니라 현실 세계에서 실제 위치를 가진다. 사용자의 위치와 방향을 서버에서 저장하기 때문에, 그림3과 같이 사용자가 대화 에이전트가 위치한 곳을 바라볼때만 에이전트가 화면에 표시되며, 사용자의 위치에 기반하여 에이전트의 크기가 달라짐으로써 원근감도 표시된다. 본 실험에서는 연구실1의 전등 스위치 옆과 연구실2의 에어컨 제어기 옆에 위치한다고 가정하였다. 사용자가 에이전트에 충분히 가깝게 접근했을 때 의사소통이 가능하며, 에이전트에게 명령을 내림으로써 현실 세계의 제어가 가능하다. 이 실험 동영상은 <https://youtu.be/nNpo-6Wtz0o> 에서 볼 수 있다.

5. 결론 및 향후 계획

본 논문에서는 현실 세계에 영향을 주는 혼합 현실 환경의 대화 시스템 구조를 제안하였다. 사용자가 클라이언트를 통해 ‘볼 커’ 등과 같은 현실세계에 영향을 주는 명령을 전달한 경우, 서버는 이 명령에 대한 적절한 시스템 응답을 생성함과 동시에 명령 수행을 위한 장치에게 필요한 동작을 수행하도록 제어한다.

본 논문에서의 대화 에이전트가 현실 세계에 영향을 줄 수 있는 기능은 총 4가지(전등 켜기, 전등 끄기, 에어컨 켜기, 에어컨 끄기)에 불과하지만, 추후 다양한 실내 공간에 gadget들을 설치하여 제어할 계획이다. 또한, 사용자가 빠르게 움직이면 가상 공간의 개체들, 이를테면 본 연구의 대화 에이전트 등의 개체들의 위치가 적절히 표시되지 않는다는 단점이 존재하는데, 이는 [5]에서 제안한 PhoenixBot과 같은 단점이며, 그 이유는 불안정한 WiFi 스캔, 불안정한 주변 자기장 등의 요소이다. 따라서, 대화 시스템의 실내위치 및 방향 인식 정확도를 지속적으로 개선할 필요가 있다. 향후 대화 에이전트가 실내 공간을 자의적으로 이동할 수 있는 인공지능에 대해 연구할 계획이다.

Acknowledgement

이 성과는 2019년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2019021348). 본 연구는 순천향대학교 연구논문진흥사업 지원으로 수행하였음 [20190755].

참고문헌

- [1] 장순필, “대화시스템의 영역 지식과 대화 전략 통합을 위한 온톨로지 모델”, 서강대학교 대학원 석사 학위논문, 2014
- [2] Amazon Alexa, <https://alexa.amazon.com>
- [3] Google Home, <https://home.google.com>
- [4] Naver Clova, <https://clova.ai/ko>
- [5] KyuHwon Park, Young-Seob Jeong, Video: Indoor

Dialog Agent in Mixed Reality, MobiSys'19, pp.708-709, 2019

- [6] Raspberrypi, https://en.wikipedia.org/wiki/Raspberry_Pi
- [7] 박영선, 안상진, 이영준, “라즈베리 파이를 활용한 프로그래밍 교육 방안”, 한국컴퓨터정보학회 동계 학술대회 논문집, 제22권, 제1호, pp.451-453, 2014
- [8] ISO8601, https://en.wikipedia.org/wiki/ISO_8601