

박사학위논문

비디오월 제어시스템을 위한
손동작과 음성 기반 멀티모달
상호작용에 관한 연구

2024년

한 성 대 학 교 대 학 원

스마트융합컨설팅학과

스마트융합제품전공

조 영 욱

박 사 학 위 논 문
지도교수 노광현

비디오월 제어시스템을 위한
손동작과 음성 기반 멀티모달
상호작용에 관한 연구

A Study on Multimodal Interaction Using Hand
Gestures and Speech for Video Wall Control
System

2024년 6월 일

한 성 대 학 교 대 학 원

스마트융합건설팅학과

스마트융합제품전공

조 영 욱

박 사 학 위 논 문
지도교수 노광현

비디오월 제어시스템을 위한
손동작과 음성 기반 멀티모달
상호작용에 관한 연구

A Study on Multimodal Interaction Using Hand
Gestures and Speech for Video Wall Control
System

위 논문을 공학 박사학위 논문으로 제출함

2024년 6월 일

한 성 대 학 교 대 학 원

스마트융합컨설팅학과

스마트융합제품전공

조 영 욱

조영욱의 공학 박사학위 논문을 인준함

2024년 6월 일

심사위원장 김 승 천 (인)

심 사 위 원 임 황 용 (인)

심 사 위 원 신 현 준 (인)

심 사 위 원 노 광 현 (인)

심 사 위 원 이 후 진 (인)

국 문 초 록

비디오월 제어시스템을 위한 손동작과 음성 기반 멀티모달 상호작용에 관한 연구

한 성 대 학 교 대 학 원
스 마 트 융 합 컨 설 팅 학 과
스 마 트 융 합 제 품 전 공
조 영 옥

디스플레이 산업 발전으로 다양한 인터페이스를 통해 컴퓨터와 상호 작용하는 부분이 증가하고 있다. 마우스나 키보드의 직접적인 제어 없이 손동작이나 음성으로 컴퓨터를 제어하거나 자동차 안에서 손동작 및 음성 명령으로 음악을 듣거나 길 안내를 할 수 있다. 음성과 손동작 인식 기술은 엔터테인먼트, 의료, 로봇틱스, 스마트시티 등에 다양한 응용 분야에 적용이 되고 있다. 자율주행자동차, 자율경비로봇, 자율배달로봇이 도시를 활보하고, 드론 택시가 하늘을 날아다니는 스마트 시티의 중요성이 커지고 있다. 도시의 다양한 문제를 모니터링하고 관제하는 스마트 통합관제센터의 기능도 고도화되고 있다. 도시뿐만 아니라 철도, 발전소, 제조, 건설 등의 안전에 대한 모니터링이 훨씬 강화되고, 중대재해법 시행으로 기업주의 관리 감독이 더욱 엄중해졌다.

본 연구는 스마트 통합관제시스템의 CCTV영상을 모니터링하고 제어가 되는 시스템인 비디오월 제어시스템과 자연 친화적인 방법으로 상호작용

하는 기능을 개발하여 관제센터 운영자가 효율적으로 모니터링하는 것을 목표로 한다. 손동작 인식하기 위해 웹카메라(RGB), 3D 카메라, Leap Motion, LiDAR 인터페이스 장비를 사용하는데 본 연구에서는 널리 보급된 웹 카메라(RGB)를 사용하였고 음성 인식을 위한 인터페이스 장치는 마이크를 활용하였다. 입력 장치에서 감지된 손동작 및 음성을 전처리 과정을 거쳐 정의된 동작 라이브러리를 활용하여 모델링을 거쳐 실제 Detection 검증 후 손동작과 음성신호를 순차적으로 동기화 및 상호작용하여 제어가 되도록 설계하였다.

멀티모달 인터페이스를 이용한 비디오월 제어시스템의 음성 및 손모양 인식에 대한 기능별 기술 요소를 구성하고, 손모양 인식은 웹캠을 통해 구글사의 머신러닝 기반 고성능 손제스처 인식 솔루션인 MediaPipe Hands를 활용하여 손동작 인식을 구현하였고, 각 손에 대해 21개의 3D 랜드마크를 프레임마다 추론하여 두 개 이상의 손을 동시에 인식하고 인터페이스가 가능하도록 구성하였다. 손모양의 인식한 후 음성 명령을 통해 비디오월의 제어판의 UI를 제어하여 영상, 이미지, 웹브라우저 등의 콘텐츠를 비디오월에 표출되고 콘텐츠를 삭제, 확대, 축소의 기능을 개발하여 마우스로 제어하는 방식에서 벗어나 자연 친화적인 비디오월 제어시스템을 구현하였다. 가장 적합한 손동작을 적용하기 위해 3가지 손동작을 테스트하여 각각 인식에 대한 정확도와 편리성을 검증하여 손동작을 선정하였다. 음성인식은 ‘MIT/ast-finetuned-speech-commands-v2’의 모델을 결합하여 음성 명령을 내리면 음성을 텍스트로 변환하여 손동작과 연동하여 동시에 제어가 되도록 설계하였다.

설계된 손동작 및 음성인식 기반 비디오월 제어시스템에 적용하여 시스템의 검증을 위해 컴퓨터공학과 학생과 ICT 종사자를 대상으로 활용 및 테스트를 하였고, 기존 마우스로 제어하는 방식과 비교하며 사용자 평가를 통해 만족도, 유용성에 대해 검증하고 평가를 진행하였다. 손동작과 음성인식으로 비디오월 제어시스템을 제어하는 기능 중에 마우스 좌클릭, 우클릭 기능이 가장 높은 90%의 유용성으로 나타났고, 마우스 이동, 더블클릭, 휠 업/다운(확대/축소)이 80% 수준으로 다소 낮게 나타났다. 그리

고 마우스 스크롤 업/다운 기능의 제어는 가장 낮은 50%대로 스크롤 바 제어의 어려움을 확인하였고, 손동작 및 음성인식으로 비디오월 제어시스템 제어판의 UI를 제어하였을 때 걸리는 시간은 손동작과 음성 인식으로 순차적으로 제어할 때보다 0.3초에서 0.55초 빠르게 동작이 인식되었다.

본 연구의 성과는 멀티모달 기반으로 비디오월 제어시스템을 손동작이나 음성으로 제어가 가능성과 유용성을 확인하였다. 그리고 손동작 인식을 하기 위해 웹캠과 같은 범용적인 디바이스에서 LiDAR, LeapMotion 등 다양한 디바이스에도 인식이 되는 플랫폼 연구 개발이 필요함을 인식하게 되었다.

또한 손동작 및 음성인식으로 명령을 내린 후 콘텐츠의 응답시간을 마우스의 기능보다 유사하거나 빠르게 콘텐츠를 제어할 수 있어야 사용성이 높다는 것을 확인하였다. 특히 제어판의 UI 중에 음향, 해상도 등을 제어하는 슬라이드 바에 대한 제어는 컴퓨터 GUI 환경을 더욱 정밀한 방법으로 제어하는 연구가 진행되어야 할 것으로 판단된다.

【주제어】 멀티모달, 비디오월 제어, 통합관제, 손동작 인식, 음성 인식

목 차

제 1 장 서 론	1
제 1 절 연구 배경	1
1) 통합관제센터의 컴퓨터간 상호작용	1
2) 인간과 컴퓨터간 상호 발전	3
3) 멀티 디스플레이 비디오월 시스템 기술 변화	5
제 2 절 연구의 필요성	8
1) 통합관제센터의 운영 환경	8
2) 인간 친화적인 인터페이스 전환	9
3) 인간 친화적인 인터페이스 확장서비스	11
4) 통합관제센터의 변화의 다양성	13
제 3 절 연구의 범위와 목적	15
1) 연구 범위	15
2) 연구 모델 기능 상세 기능	16
3) 연구 모델 설계와 기술 요소	17
4) 연구 목적	19
제 2 장 이론적 배경 및 선행연구	20
제 1 절 손동작 인식을 활용한 상호작용	20
1) 손동작 인식 개요	20
2) 손동작 인식에 대한 기존 연구	21
3) 컴퓨터를 제어의 손동작 형태	24
4) 손동작 인식 활용사례	27
제 2 절 음성 인식에 활용한 상호작용에 관한 연구	30
1) 음성 인식의 개요	30
2) 음성 인식에 대한 기술	31
3) 음성 인식의 산업 활용 사례	35
제 3 절 멀티모달을 활용한 상호작용 연구	38
1) 멀티모달 시스템 개요	38

2) 멀티모달 시스템 활용 사례	41
제 3 장 멀티모달 기반 비디오월 제어시스템 상호작용 설계	44
제 1 절 멀티모달의 비디오월 제어의 상호작용 정의	44
1) 비디오월 일반적인 기능 정의	44
2) 비디오월 제어의 정의	45
3) 멀티모달 인터페이스의 기능적 정의	45
제 2 절 멀티모달 기반 비디오월 제어시스템 아키텍처	47
1) 비디오월 제어시스템의 상호작용 체계	47
2) 입력 모달리티 전처리 및 모델링	48
3) 입력 동기화와 인터랙션 매니저	49
제 3 절 멀티모달 기반 비디오월 시스템 설계	50
1) 손동작 및 음성 인식 비디오월 제어시스템 기능 정의	50
2) 손동작 인식 기반 상호작용 설계	55
3) 음성 인식 기반 상호작용 설계	66
4) 손동작과 음성 인식 기반 상호작용 설계	70
제 4 장 구현 및 테스트	74
제 1 절 구현 범위	74
1) 제어시스템 범위	74
2) 제어시스템 인프라 구성	75
제 2 절 시스템 성능 구현 및 사용성 평가	77
1) 성능 구현 및 평가	77
2) 사용자 평가	87
3) 연구 성과	93
제 5 장 결 론	94
참 고 문 헌	96
ABSTRACT	110

표 목 차

[표 1-1] 내추럴 사용자 인터페이스 입출력 방식	11
[표 1-2] 연구모델 상세기능	16
[표 1-3] 연구 모델 기능별 기술 요소	18
[표 2-1] 음성 인식 발전 과정	30
[표 2-2] 음성 인식 시스템 적용	31
[표 2-3] 음성 인식 기술 활용 국내외 사례	36
[표 2-4] 대규모 언어모델(LLM) 주요 서비스	37
[표 2-5] 사람과 기계간 입출력 모달리티	38
[표 2-6] 멀티모달 인터페이스 종류	39
[표 2-7] 손동작 및 음성 인식 적용 가능성 비교	40
[표 3-1] 주요 인터페이스 기능	46
[표 3-2] H사의 비디오월 제어판 UI 화면 기능 및 적용 동작	52
[표 3-3] 손동작 및 음성 인식 기반 비디오월 제어판 기능 정의	53
[표 4-1] 손동작 및 음성 인식 비디오월 제어별 기능	75
[표 4-2] 손동작 및 음성 인식 위한 하드웨어 및 소프트웨어 구조	76
[표 4-3] 3가지 손동작 구현 테스트 결과 1	77
[표 4-4] 3가지 손동작 구현 테스트 결과 2	78
[표 4-5] 3가지 손동작 구현 테스트 결과 3	80
[표 4-6] 3가지 손동작 구현 테스트 결과 4	83
[표 4-7] 3가지 손동작 구현 테스트 결과 5	84
[표 4-8] 손동작 및 음성인식 기능 정의 테스트 결과	86
[표 4-9] 사용자 평가 구성 정보	88
[표 4-10] 사용자들이 평소 컴퓨터 상호작용 대한 평가	89
[표 4-11] 손동작 및 음성 인식 시스템의 기능 구현 만족도 평가	90
[표 4-12] 손동작과 음성 인식 동시 구현 만족도 평가	91
[표 4-13] 손동작과 음성 인식 동시 구현 시간 만족도 평가	92

그 림 목 차

[그림 1-1] CCTV 통합관제 디스플레이	2
[그림 1-2] 멀티모달 개념	3
[그림 1-3] 확장 현실 기반 몰입형 CCTV 관제센터 개념도	6
[그림 1-4] 스마트월(Smart-Wall) 개념도	7
[그림 1-5] 사용자 인터페이스 발전 방향	10
[그림 1-6] KT 음성 인식 서비스 기기	12
[그림 1-7] 인공지능을 활용한 디지털 통합관제센터	14
[그림 1-8] 연구 모델 시스템 흐름도	17
[그림 2-1] 손동작 및 마우스 이벤트	27
[그림 2-2] logbar사의 스마트 반지와 윈도우 기반 3D핸드 트래커	28
[그림 2-3] 애플사 비전 프로 동작 장면	29
[그림 2-4] 음성 인식 과정	32
[그림 2-5] 카 하는 글자를 발생시 음성 파형	32
[그림 2-6] 모델별 단어 수준 오류율	33
[그림 2-7] 연도별 음성 인식 모델	34
[그림 2-8] 멀티모달 기반 스마트홈 AI 에이전트	41
[그림 2-9] 멀티모달 센서 경비 로봇	42
[그림 2-10] 멀티모달 융합 사례	43
[그림 3-1] 멀티모달 기반 비디오월 제어 아키텍처	47
[그림 3-2] H사의 비디오월 제어시스템 UI화면	50
[그림 3-3] H사의 비디오월 ‘A’ 제어판 UI화면	51
[그림 3-4] 제어판과 제어스크린 기능별 표출 콘텐츠	54
[그림 3-5] MediaPipe Hands Landmarks	56
[그림 3-6] MediaPipe 활용 손동작 인식 Workflow	57
[그림 3-7] 손동작 리스트	58
[그림 3-8] 마우스 이동 손동작	59

[그림 3-9] 마우스 더블클릭 손동작	60
[그림 3-10] 마우스 좌클릭 손동작	60
[그림 3-11] 마우스 우클릭 손동작	61
[그림 3-12] 마우스 스크롤다운 손동작	61
[그림 3-13] 마우스 스크롤업 손동작	62
[그림 3-14] 마우스 휠업 손동작	62
[그림 3-15] 마우스 휠다운 손동작	63
[그림 3-16] 마우스 포인트로 이미지 및 영상 이동 손동작 1	64
[그림 3-17] 마우스 포인트로 이미지 및 영상 이동 손동작 2	65
[그림 3-18] 마우스 포인트로 이미지 및 영상 이동 손동작 3	66
[그림 3-19] 오디오 스펙트럼 트랜스포머 아키텍처	67
[그림 3-20] 음성 인식 활용한 화면 제어1	68
[그림 3-21] 음성 인식 활용한 화면 제어2	69
[그림 3-22] 손동작과 음성 인식 활용한 제어시스템 1	71
[그림 3-23] 손동작과 음성 인식 활용한 제어시스템 2	72
[그림 4-1] 손동작 및 음성 인식으로 비디오월 제어 구조	74
[그림 4-2] 손동작 및 음성 명령 콘텐츠 표출	79
[그림 4-3] 손동작으로 콘텐츠 이동하여 음성으로 right 명령	81
[그림 4-4] 손동작으로 콘텐츠 이동하여 음성으로 left 명령	81
[그림 4-5] 손동작 및 음성 명령 제어판 스크롤바 아래로 제어	83
[그림 4-6] 손동작 및 음성 명령 제어판 스크롤바 위로 제어	83
[그림 4-7] 손동작과 음성 명령 콘텐츠 확대	85
[그림 4-8] 손동작과 음성 명령 콘텐츠 축소	85
[그림 4-9] 손동작 및 음성으로 비디오월 제어로 콘텐츠 표출	87

제 1 장 서 론

제 1 절 연구 배경

1) 통합 관제센터의 컴퓨터간 상호 작용

디스플레이 장치의 급속한 발전으로 다양한 새로운 인터페이스의 수요가 증가하고 있고, 인간과 컴퓨터 간 상호작용을 위해 마우스나 키보드의 직접적인 제어 없이 음성이나 동작으로 공간을 터치하는 단순하고 직관적인 제어방법으로 연구가 증가하는 추세이다.¹⁾ 음성 인식이나 동작 인식을 이용한 휴먼 인터페이스(Human Interface)의 발달은 엔터테인먼트, 의료, 로봇틱스, 스마트 시티 등 다양한 분야에 응용되고 있으며,²⁾³⁾ 사람들이 학습 없이 쉽게 자연스러움(손동작, 음성, 눈길 등)으로 스마트 기기를 조작하고 복잡한 컴퓨터 기능을 단순하게 동작 및 제어함으로써 운영의 편리성을 극대화하고자 하는 움직임이 증가하고 있다.⁴⁾

2020년부터 2022년간 팬데믹의 여파로 방송사 및 엔터테인먼트사들이 오프라인 공연을 랜선 공연으로 전환 및 활성화되어 다양한 프로그램을 제작하였고, 새로운 방송 및 공연 문화를 만들어지면서 비대면 방송 기술의 비약적으로 발전하여 운영자가 더 편리하게 운영할 수 있도록 운영자 프로그램이 개발되고, 대형 디스플레이에 영상을 표출하는 방식이 다양한 방법으로 연구되고 있다.⁵⁾

1) 이광형, 신동규 and 신동일. (2014). 직관적인 핸드 모션에 기반한 NUI/NUX 프레임워크. 『인터넷정보학회논문지』, 15(3), 11-19.

2) I. Oikonomidis, N. Kyriazis, and A.A.(2011) Argyros, "Efficient Model-based 3D Tracking of Hand Articulations using Kinect," Proc. of British Machine Vision Conf., pp. 101.1-101.11.

3) 노정규. (2014). LCD Display 설비 Contents의 Kinect기반 동작제어 기술 구현에 관한 연구. 『전기학회논문지』, 63(4), 565-569.

4) 유인호, 신동윤. (2012). Easyoloy 학습조차 필요없는 편안함 추구해야. 『동아비즈니스리뷰』.

5) 유건식. (2020). 코로나19 시대와 방송 제작 방식의 변화. 『방송과미디어』, 25(4), 51-60.

또한 지능적인 사물과 결합하여 사용자와 직접적인 소통의 도구인 DID (Digital Information Display) 방식에서 LED 기술의 발달로 실내외에 LED 디스플레이 방식으로 영상을 표출하는 형태로 발전하고 있고, 사용자와 상호작용하는 다양하고 자연스러운 인터페이스의 필요성이 대두되고 있다.⁶⁾

범용적으로 사용하는 PC나 서버의 모니터를 통해 대형 디스플레이 장치에 표출되는 이미지나 영상을 제어하는 방식은 전통적인 키보드, 마우스로 조작하여 UI를 클릭하여 대형 디스플레이 표출이 되어 간단한 조작으로 입력하는 경우 별도의 제어장치가 필요하다.⁷⁾



[그림 1-1] CCTV 통합관제 디스플레이

전국 시군구 지자체 226개 통합센터에서 운영하는 CCTV 영상을 관제사가 [그림 1-1]과 같이 키보드나 마우스로 방법, 쓰레기 투기 방지, 주차관리, 주/정차 단속, 재난화재 감시, 시설물 관리 등의 영상을 검색하거나 제어하고 있다.⁸⁾

6) 김정수. (2017). DID 복합 콘텐츠 제어를 위한 키넥트 기반 인터페이스. 『대구 : 한국교육학술정보원』

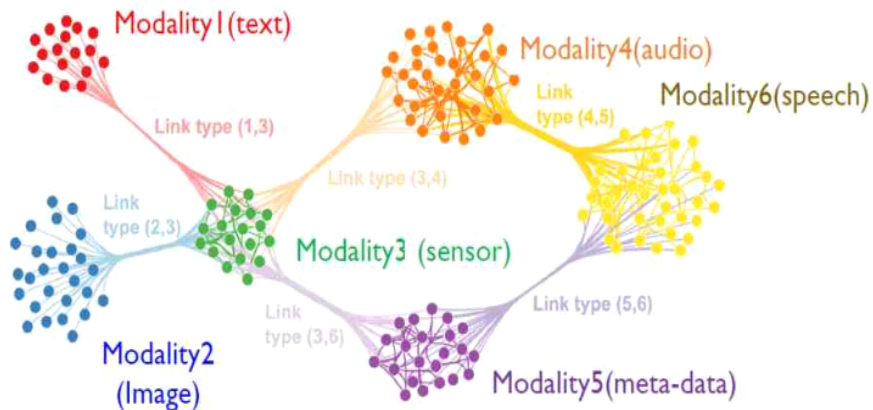
7) 이민규, 전재봉 (2012). 키넥트를 이용한 개인용 컴퓨터 제어. 『한국정보과학회 학술발표논문집』, 39(1A), 343-345.

8) 정충식. (2015). CCTV 통합관제센터의 운영 개선 방안에 관한 연구: 부산시 통합관제센터를 중

관제센터의 운영자가 마우스가 키보드로 대형 DID 시스템이나 LED 디스플레이를 제어하는 방법에서 인간과 컴퓨터 간 상호작용하는 손동작이나 음성 등으로 인간이 자연스러운 동작으로 제어하는 대화형 통합관제 시스템의 연구가 활발하게 진행되고 있다.⁹⁾ 최첨단 스마트 시티의 진화는 점점 늘어가고 있는 상황에서 통합관제 시스템 운영에 대한 부분도 기존의 방식보다 인간 친화적인 방식으로 개선되고 연구되어야 한다.

2) 인간과 컴퓨터간 상호작용 발전

인간과 컴퓨터간 상호작용에서 멀티모달 인터페이스 방식으로 인간의 자연스러운 행동(제스처, 손동작, 음성, 행동 등)들에 대한 정보를 해석하는 인지 기반 기술로 상호작용하는 연구가 진행되었다.¹⁰⁾ 다양한 입력 요소를 통해 상호작용하는 인간 중심형 인터페이스 기술인 멀티모달 인터페이스 기술이 지속적으로 발전하고 있다.¹¹⁾



[그림 1-2 멀티모달 개념]

심으로. 『한국지역정보학회지』, 18(3), 123-154.

9) Zahra, R., Shehzadi, A., Sharif, M. I., Karim, A., Azam, S., De Boer, F., ... & Mehmood, M. (2023). Camera-based interactive wall display using hand gesture recognition. *Intelligent Systems with Applications*, 19, 200262.

10) 임미정. (2006). 멀티모달 인터페이스 개발을 위한 휴먼-컴퓨터 인터랙션 설계. 『아주대학교』

11) 조윤상, 이지윤, 이영재, 강현규, 김성범.(2019).Multimodal Deep Learning for Explainable Product Design Data. 『대한산업공학회 추계학술대회 논문집』, 2019(11) 1184-1216.

모든 상호작용의 본질은 멀티모달이며,¹²⁾ 멀티모달은 전통적인 텍스트 외에 음성, 동작, 표정, 시선, 생체 등 여러 가지 입력방식을 융합하여 인간과 컴퓨터 간에 자연스러운 의사소통이 가능한 사용자 친화형(User-Friendly)기술이다.¹³⁾

음성이나 동작 등의 인간 친화적인 모달리티를 사용하여 컴퓨터와 인터페이스를 자연스럽게 하는 멀티모달 연구이며, 유니모달과 비교할 때 사용이 편리성과 안정성이 있으며, 새롭고 빠른 기능이 강조되고 있다.¹⁴⁾ 기존의 텍스트, 이미지, 음성 등의 유니 모달 방식의 연구에서 음성, 제스처, 시선, 표정, 생체 신호 등 다양한 입력을 기반으로 하는 인터페이스 기술들이 연구되고 있다.¹⁵⁾

IT 산업의 발전과 더불어 컴퓨터 인터페이스에 대한 상호작용 중에 손동작 제스처 기반의 인터페이스 방식들이 손 모양이나 동작을 기호화하여 활용되고 있다.¹⁶⁾ 컴퓨터 활용의 보편화로 마우스의 지속적인 사용이 손목터널증후군 등의 직업병 발생이 빈번하게 보고되고 있다.¹⁷⁾

이러한 문제를 해결하기 위해 멀티모달 입력 장치가 마우스 포인트의 위치를 계산하고 얼굴 영상 인식 스프레드와 다른 영역으로 이동하기 위해 입력하는 해당 영역의 번호를 음성 인식으로 처리하는 장치도 연구되고 있다.¹⁸⁾ 그리고 음성 인식 분야에서 여러 가지 입력 정보, 제스처, 얼굴, 손짓, 눈동자의 움직임을 연동하여 정보를 정확하게 인식하고 기존의 시스템의 성능을 개선하는 연구도 활발하게 움직이고 있다.¹⁹⁾

12) Norris, S. (2006). Multiparty interaction: a multimodal perspective on relevance. *Discourse Studies*, 8(3), 401–421.

13) [http://blog.naver.com/PostView.naver?blogId=yckim002&logNo=222599463531
& parentCategoryNo=&categoryNo=63](http://blog.naver.com/PostView.naver?blogId=yckim002&logNo=222599463531&parentCategoryNo=&categoryNo=63)

14) 이동우, 손용기, 김지은, 선우준, 조일연, 한동원. (2007). 웨어러블 컴퓨터를 위한 음성 제스처 기반 멀티모달 융합 사용자 인터페이스. 『한국차세대컴퓨팅학회』, 3(2), 17–25.

15) Vinciarelli, Alessandro(2009). Social signal processing: Survey of an emerging domain. In *Image and Vision Computing*, 27(12), 1743–1759

16) 이창주, 이준호, (2007) “깊이 에지 기반의 Curvature Scale Space Map을 이용한 손 제스처 인식”, 『한국정보처리학회』 제14권 제1호

17) Bhandari, Dinesh J. Mishra, Daxa G. Parikh, Shweta M. Sharma, Deepak B. (2017). Computer Use and Carpal Tunnel Syndrome: A Case-control Study. *Indian Journal of Occupational & Environmental Medicine*. Vol. 21 Issue 3, p109–114. 6p. 6 Charts.

18) 최원석, 이동우, 김문식, 나종화.(2007).음성인식 및 영상처리 기반 멀티모달 입력장치의 설계. 『제어로봇시스템학회』, 13(8),743–748.

멀티모달 인터페이스 기술 기반의 멀티모달 이미지 시장은 Globenewswire(2021년) Multimodal Imaging Global Market Trajectory & Analytics 분석에 의하면 2020년 230억 달러에서 연평균 4.4% 성장하여 2027년에는 310억 달러로 성장할 것으로 예상되며 이 분야의 기술은 꾸준히 발전할 것으로 보인다.

3) 멀티 디스플레이의 비디오월 기술 변화

최근 멀티 디스플레이를 활용한 생활 정보 서비스와 다양한 콘텐츠를 제공하며 일상의 변화를 주는 역할로 발전하고 있고, 공공장소는 물론 상업공간, 옥외공간 어디서나 볼 수 있는 생활 밀착형 매체로 자리 잡아 생활 전반에 확대되고 있는 멀티 디스플레이 인터페이스 연구가 대두되고 있다.²⁰⁾

비디오월이란 여러 대의 컴퓨터 모니터, 비디오 프로젝터 또는 TV를 나란히 놓거나 쌓아 하나의 대형 화면을 형성하는 특수한 다중 모니터 설정이다. 일반적인 디스플레이 기술에는 LCD 패널, DvLED 어레이, 혼합 프로젝션 스크린, 레이저 인광체 디스플레이 및 후면 프로젝션 큐브가 포함된다.²¹⁾

비디오월은 스마트 시티 구축과 함께 지능형 관제시스템의 중요성이 점점 커지고 있으며, 모니터링 방법도 점점 더 정교하고 지능화되고 있다. 상황 인지 기술과 멀티모달 기술을 활용하여 “언제, 어디서든” 상황을 객관적이고 정확하게 파악하고, 적절한 서비스 방법을 추론하여 멀티모달 인터랙션 서비스를 제공하는 관제시스템에 적용되고 연구되었다.²²⁾

스마트 시티의 차원을 넘어 스마트월드의 여러 가지 형태 스마트 산업, 스마트 환경, 스마트 그리드, 스마트 홈, 스마트 농업 등과 같이 기존의 생활 방식을 바꾸고 있다. 결국에는 사이버, 물리적인, 사회적, 그리고 생각의 공간

19) 최광국, 곽상훈, 김유진, 김철, & 최승호. (2000). 멀티모달을 이용한 응용프로그램 제어에 관한 연구. 『한국음향학회』, 205-208.

20) 정유경. (2011). 멀티 디스플레이 사례분석과 관리 인터페이스에 관한 연구. 디자인지식저널, 18, 180-189.

21) <https://academic-accelerator.com/encyclopedia/kr/video-wall>

22) 박성도,장복만,장효경,김봉희,최의인.(2010) 멀티모달 인터랙션을 이용한 관제시스템 설계. 『한국통신학회』 : 360-361.

들까지 확대되어 스마트 관제센터의 모니터링은 더욱 증대되고 있다.²³⁾

물리적인 스마트 월드에서 가상에서 동일하게 적용하여 구현하는 모델링인 디지털 트윈(Digital Twin) 개념의 전환되어 실제 사물과 연동하여 모니터링, 분석, 예측 기법을 통해 의사결정에 활용하고 실제적인 물리적 공간과 가상 디지털 공간을 상호 작용하는 기술로 발전하고 있다.²⁴⁾

스마트 시티의 지능형 영상 관제 기술의 발전은 메타버스 기반 첨단 도시 통합 관제플랫폼 구축·운영 필요성에도 대두되고 있다. 확장 현실 기반 몰입형 관제 기술을 메타버스 공간에 적용하여 시·공간의 제약을 탈피하고, 시인성이 높은 모니터링 가능하고 신속한 의사결정, 혼합 현실 기반 원격관제 및 원격지원, 실시간 스마트 정보 알림 기술 등의 솔루션을 적용하여 신속하게 실시간 대응이 가능하다.²⁵⁾ [그림 1-3]은 확장 현실 기반 몰입형 CCTV 관제기술 개념도이다.



[그림1-3] 확장 현실 기반 몰입형 CCTV 관제센터 개념도

23) 한충희, 한창희. (2021). 안전한 스마트월드를 위한 사이버 테러위협 제거 방안 연구. 『융합보안논문지』, 21(1), 107-113.

24) 여성수, 김재원, 이재용. (2021). 도시 환경에서의 디지털트윈 물리보안 시스템 논의. 『한국경찰학회』, 23(6), 153-177.

25) 김준철. (2021). 확장현실 기반 몰입형 CCTV 스마트 관제 기술. 『서울기술원』, 21(8), 39-41.

확장 현실 기반의 몰입형 관제센터는 플랫폼으로서 스마트 도시에 부각되고 있다. 스마트 시티의 핵심기술인 빅데이터 클라우드, 인공지능, 사물인터넷, 가상 현실등의 접목되어 초연결·초지능을 가속화시키고 있으며, 모빌리티와 통신망이 통합·연계되는 새로운 디지털 스마트 시티의 탄생과 지능형 플랫폼의 핵심이다.²⁶⁾ 이처럼 통합운영센터의 운영형태가 다양한 방식으로 변화되고 있는 것을 알 수 있으며 운영방식의 진화는 계속 진행될 것이다.

미래 관제센터는 초실감 상호 기술을 이용한 원격의 다지점(Multi-point)의 공간을 초실감으로 연결할 수 있는 스마트 월(Smart-Wall)기술에도 적용되며, 컨트롤러나 버튼 없이 사용자의 움직임이 그대로 컴퓨터와 상호 작용하는 수단도 되고, 음성, 동작, 표정, 터치, 텍스트 등 인간 친화적인 다차원 입력 인식 기술, 다자간 체감 인터랙션 기술 등으로 발전되고 있다.²⁷⁾



[그림 1-4] 스마트월(Smart-Wall) 개념도

[그림 1-4]는 초실감 상호작용 기술을 적용한 스마트 월 개념도를 보여준다. 비디오월은 초실감 상호작용 기술과 접목되어 가상공간보다 진보된 개념으로 가상과 현실을 상호작용하여 그 속에서 사회·경제·문화 활동이 이루어지면서 가치를 창출하는 요소 기술로 활용될 것이다.²⁸⁾

26) 김덕인,서종국,(2022) 광역 스마트도시 안전망 구축에 관한 연구. 『인천대학교』

27) 김도영, 강진아, 이미숙 (2015). 스마트 월(Smart-Wall)- 원격 다지점 공간의 초실감 가상연결 기술. 『전자통신동향분석』, 32(2), 34-40.

28) 이하은, 한정협. (2021). 메타버스 플랫폼의 체험형태에 따른 유형 분류 및 특성 연구:실감, 초

제 2 절 연구의 필요성

1) 통합관제센터 운영 환경

전국 현재 228개 시군구 중 179개에서 상황실을 운영하지 않고 있고, 일부 시군구 123개만 통합관제센터의 영상을 경찰과 소방이 공동으로 활용하는 상황이고, 구축된 CCTV 화질 저하, 운영 요원의 인식 부족 등으로 종합적인 운영의 어려움이 발생하여 긴급 안전사고가 발생하면 관계 기관과 유기적으로 대응을 하지 못해 이태원 참사와 같은 대형 참사로 이어진 것이다.²⁹⁾

또한 통합관제센터의 운영 요원의 1인당 관리하는 CCTV 수량이 200대 이상을 관리하는 부분도 이러한 문제를 더욱 증가시키고, 예산 부족으로 인한 실질적인 통합관제센터 운영이 어려움이 발생하고 있다. 이러한 문제를 해결하기 위해 첨단 스마트 시티 통합관제센터 구축, 빅데이터 활용, 실감형 상호작용기술을 활용한 능동적인 운영이 가능한 기능으로 고도화할 필요가 있는 것이다.³⁰⁾

미래의 도시의 공간에서는 자율주행 자동차와 드론이 일반화되고 자율형 경비로봇이 거리를 활보하고, 다양한 IT 기기와 상호 소통하는 시대로 발전할 것으로 전망된다. 따라서 스마트 통합관제센터의 모니터링 기능은 공상과학영화에서 나오는 스마트 월과 대화하며 다양한 제스처로 명령을 내리고 모니터링하는 기능으로 발전할 것이다. 현재는 실감형 상호작용 기술이 선별적으로 일부 서비스에만 활용되지만, 점진적으로 기술의 발전이 되고 서비스 범위도 확대될 것이다.

본 연구에서는 스마트 시티의 스마트 통합관제센터의 멀티모달 기반의 통합관제센터의 제어하는 부분에서 손동작과 음성 인식을 중심으로 운영 요원들이 컴퓨터의 마우스와 키보드 없이 제어할 수 있는지 사용성과 유용성을 연구하였다.

실감 메타버스 대표사례를 중심으로. 『한국공간디자인학회』, 16(8), 427.

29) <https://www.asiatoday.co.kr/view.php?key=20230419010011474>

30) 장일식, 박종철. (2018). CCTV 통합관제센터의 운영성과 및 개선 방안-구미시를 중심으로 -. 『경찰학논총』, 13(1), 209-246.

기존의 음성 인식과 제스처 인식 기술은 다양한 분야에서 연구되어 왔고 유니모달에서 멀티모달로 연구는 더욱 가속화되고 있으며, 현재 인간이 사용하고 있는 모든 IT 기기들은 인간 친화적인 기능으로 전환되고 서비스될 것으로 전망된다.

더군다나 고령화의 가속화로 인간 친화적인 헬스케어 기기의 발전과 IT 기기에 대한 접근성이 높은 제품이 인기를 누리고 있다. 컴퓨터와 인간이 공존하는 사회는 오래되었지만 대부분 사람들은 아직도 컴퓨터에 대해 사용이 서툴고 어려워한다. 글을 쓰려면 자판의 기능을 연습해야 하고, 마우스를 사용하여 검색할 줄 알아야 한다.

그러나 이런 기능을 몰라도 즉 음성이나 제스처로 Chat GPT와 대화를 하고 그림을 그리고 코딩도 해주고, 다양한 IT 기기를 제어도 하는 시스템들이 점점 더 증가하고 있다.

2) 인간 친화적인 인터페이스 전환

마우스, 키보드 없이 컴퓨터에 명령을 내릴 수 없을까? 라는 이런 문제의식에서 탄생한 것이 NUI(Natural User Interface)이다.³¹⁾ 화면에 어떤 표시를 이용해서 사람이 어떻게 의사를 전달(클릭, 키보드 입력 등)할지 마련된 인터페이스이다.

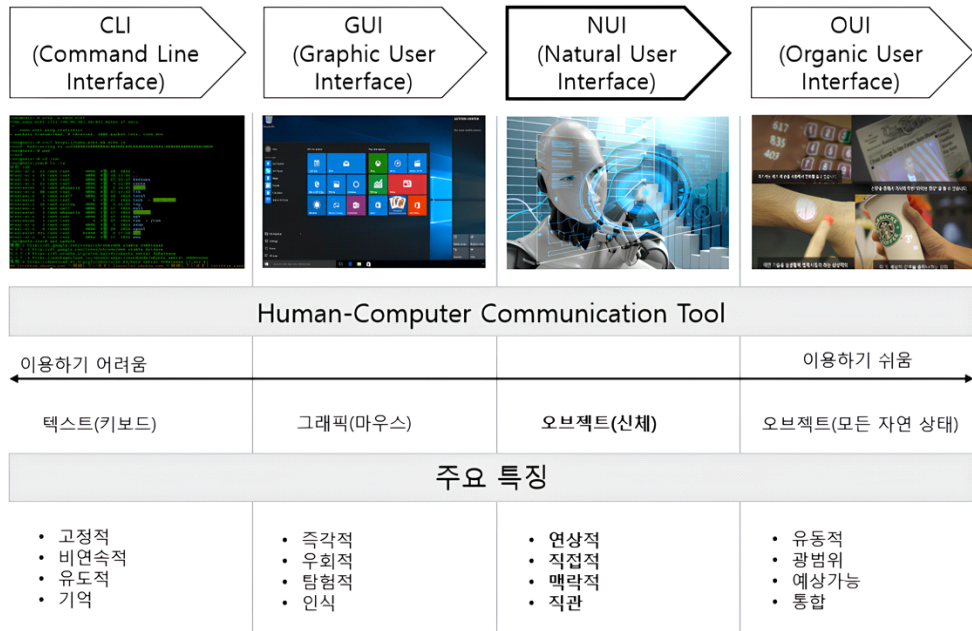
어떻게 소통하는지 따라 CLI, GUI, NUI는 용어로 나누며, CLI(Command-Line Interface)는 글자의 입출력을 통해 사용자와 컴퓨터 간 소통하는 방식이며, 대표적인 화면은 윈도우 cmd 창이고, GUI(Graphical User Interface)는 그래픽을 이용해 사용자와 소통하는 방식이며, 그래픽으로 전달되는 그림이나, 아이콘을 통해서 소통하고, 마우스로 화면에 나오는 그림을 끌어다 놓거나(Drag and Drop), 클릭해서 컴퓨터와 소통한다.³²⁾

NUI(Natural User Interface)는 사용자가 새로운 명령어와 사용법을 배우지 않더라도 실제 행동 그대로 키보드나 마우스 없이 신체 부위를 이용하여 자연스럽게 사용할 수 있다는 점이 가장 큰 장점이며, 별도로 외우거나 학습해야 하는 부분

31) https://dbr.donga.com/article/view/1101/article_no/4990

32) <https://sorrow16.tistory.com/209>

이 없고, 일상생활에서 자연스럽게 취하는 동작을 기반으로 손가락이나 제스처, 음성 등을 통해 기기를 제어할 수 있으며, 직관적인 사용이 가능한 사용자 인터페이스이다.³³⁾ [그림 1-5]은 내추럴 사용자 인터페이스 발전 방향을 보여 준다.



[그림 1-5] 사용자 인터페이스 발전 방향

내추럴 사용자 인터페이스의 인터랙션 기술은 입출력 방식에 따라 신체가 직접 기기에 접촉하는 ‘접촉식 방식’과 신체에 직접 접촉하지 않는 ‘비접촉식 방식’으로 크게 나눌 수 있으며, 다양한 방식으로 지속적인 발전을 거듭하고 있다.³⁴⁾

[표 1-1]은 내추럴 사용자 인터페이스 입출력 방식에 따라 상세기능을 보여주고 있으며 입력 기술에서 컴퓨터 비전, 음성인식, 터치, 다양한 센서, 뇌파 등의 기술이 있으며, 출력 부분의 기술은 청각, 촉각, 시각으로 인터페이스를 하는 것을 볼 수 있다. 다양한 입출력의 인터페이스 기술인 멀티모달 방식으로 인터페이스 기술은 많은 발전과 시도를 통해 기술적인 한계를 극복할 가능성이 크다.

33) 이대점, 김원, 윤여경.(2020). NUI를 활용한 가상현실 박물관 전시 디자인 연구. 『한국상품문화 디자인학회』, 62,99-109.

34) 김진형, 김현.(2015).사용자 경험 중심의 내추럴 인터랙션 디자인 방향 연구. 『한국디자인문화 학회』, 21(2), 175-186.

[표 1-1] 내추럴 사용자 인터페이스 입출력 방식

입출력 기술		상세기능
입력	컴퓨터 비전	얼굴인식, 공간인식, 증강현실, 제스처(동작)인식
	음성 인식	음성 명령, 자연어, 환경 소음
	터치	단일 터치, 멀티 터치, 펜 입력
	센서	GPS, 가속도계, 주변밝기, 생체인식(지문, 홍채 등)
	뇌파	마인드 컨트롤, 감정 인식
출력	청각	음성합성, 음성피드백
	시각	그래픽, 시각적 피드백, 가상 아바타
	촉각	감각 피드백(진동, 신호등), 촉각 피드백(감촉, 재질)

3) 인간 친화적인 인터페이스 확장 서비스

비대면(언택트) 디지털 경제의 가속화로 실감형 콘텐츠에 관심이 높아졌고 사용자의 관점에서 감각과 인지를 유발하여 실제와 유사한 경험 및 감성을 확장하는 상호작용 기술로 몰입감과 다양한 경험을 제공하는 측면에서 기존과는 다른 형태 서비스이다.³⁵⁾

실감형 콘텐츠를 위한 서비스는 사용자가 기존 범용입력장치인 키보드와 마우스로 콘텐츠를 조작하는 것이 아니라 실감형 인터페이스로 디지털 콘텐츠와 상호 작용한다.³⁶⁾ 또한 가상현실 VR(Virtual Reality), 증강현실 AR(Augmented Reality), 확장현실 XR(Extended Reality) 기술이 5G 통신 기술과 접목되면서 사회 전반적인 분야에서 메타버스(Metaverse)라는 가상세계(Virtual Worlds)를 통해 제조, 건설, 공공, 교육, 방송, 엔터테인먼트 등 분야에서 활용되고 있다.³⁷⁾

실감형 콘텐츠 제작하기 위해서는 3차원 공간 정보(depthmap)를 생성해내는 센싱(sensing) 기술, 콘텐츠와 상호작용을 위한 동작인식(gesture recognition) 기술, 3D 그래픽 구현을 위한 실시간 렌더링(real time

35) 성민정. (2020). 실감형 인터랙션 미디어 환경의 경험적 가치향상을 위한 프레즌스 요소에 관한 연구. 『한양대학교 대학원』, 서울

36) 김준우, 맹준희, 주지영, 임광혁.(2010).멀티 터치스크린과 실감형 인터페이스를 적용한 과학 실험 학습 시스템. 『한국콘텐츠학회』, 10(8),461-471.

37) 정석용, 박형수, 강환수, 조진형, Larry S Kim.(2023).메타버스 기반 실감 교육용 플랫폼 개발. 『한국컴퓨터정보학회』, 28(2),77-87. 재인용

rendering) 및 저작도구 기능이 요구되며, 마이크로소프트사의 키넥트(Kinect) 카메라를 이용한 동작인식 게임 및 증강현실 시스템이나 LeapMotion으로 새로운 인터페이스로 변화해 가고 있다.³⁸⁾

자비스(Just A Rather Very Intelligent System, J.A.R.V.I.S)는 그냥 더 똑똑한 시스템의 약어로 아이언 맨의 주인공인 토니 스파크가 가진 많은 인공지능 비서 중 하나이며, 이러한 시스템은 현실에서 퍼셉추얼 컴퓨팅(perceptual computing)이라고 하며, 용어가 어려운 것 같지만 그냥 컴퓨터 기기를 마우스나 키보드 기기 대신 음성으로 컨트롤이 가능한 것이다.³⁹⁾

국내 통신회사 KT의 기가 지니나 애플사의 아이폰의 시리, 삼성전자 갤럭시 폰의 빅스비 같은 것이며, 조금 더 정확하게 이야기하자면 퍼셉추얼(Perceptual) 컴퓨팅은 음성뿐만 아니라 사람의 손짓, 발짓 그리고 몸짓, 목소리 감정까지 알아차리고 알맞게 동작하는 기술이며, 영화상에서도 주인님이 우울해 보이면 자비스가 말동무해 주고 주인이 무엇이 필요한지 자기가 먼저 권유한다.⁴⁰⁾



[그림 1-6] KT 음성인식 서비스 기기

[그림1-6]은 음성으로 제어한 IT 기기들이며 사람의 감정을 이해하지는 못하지만 대응하는 시스템이며, 더욱더 연구해야 하는 과제이지만 캐나다의 스타트업 기업의 포커스 스마트 글라스 같은 경우엔 인공지능 비서와 연계하여 홀로그램을 제어하는 기술을 상용화까지 하였고, 자비스와

38) 김명하, 홍현기.(2013).실시간 동작인식 및 렌더링 기술을 이용한 체감형 콘텐츠 저작도구 개발. 『디지털디자인학연구』,13(3),657-665.

39) <http://www.mind-journal.com/news/articleView.html?idxno=1326>

40) <https://www.greened.kr/news/articleView.html?idxno=311733>

같은 퍼셉추얼(Perceptual) AI 시스템이 어느 정도 현실에 사용하고 있는 서비스이다.⁴¹⁾

이처럼 가상현실 세계와 상호작용하고 실감성과 몰입감을 높이는 인터페이스 기술로 미디어, 엔터테인먼트, 교육, 제조, 의료, 공공 등 다양한 분야에서 응용되고 있다.⁴²⁾

4) 통합관제센터의 변화의 다양성

디지털 기반의 통합관제 플랫폼은 IoT 인식 기술을 기반으로 전기 기계 시스템, 지능형 시스템 등 모니터링 대상 전반의 데이터를 통합해서 인구 흐름, 차량 통행, 에너지 소비량, 활동 데이터 등 다양한 지표를 한눈에 파악할 수 있으며, 디지털 트윈 시설을 통한 다양한 상황 시뮬레이션이 가능, 통합 데이터를 활용해서 좀 더 효율적인 모니터링이 가능한 시스템이다.⁴³⁾

통합관제 시스템은 특정 시설물에 대한 방법, 보안, 범죄예방 등을 목적으로 CCTV 영상을 시스템을 설치하여 통합적으로 모니터해 왔고, 점점 유무선 통신 기술의 발달로 사용자는 언제 어디서든지 디지털화된 영상을 모니터링이 가능한 시스템이다.⁴⁴⁾ 스마트 시티 및 인공지능을 활용한 통합관제 플랫폼은 도시와 대형 건물을 인공지능, 빅데이터, 디지털 트윈, 5G 특화망을 이용한 데이터의 수집, 설비 관리 제어를 한 번에 할 수 있는 DX 신기술을 적용되고, 건물, 모빌리티 등 다양한 분야에서 고객 경험을 혁신하고 있다.⁴⁵⁾

[그림1-7]은 인공지능을 활용한 디지털 통합관제센터를 보여주고 있으며,⁴⁶⁾빌딩이나 도시의 다양한 모니터링 대상을 최적화하여 관리하고 제어하는 통합센터로 구성되어, 운영자의 마우스가 키보드의 작동 없이 인간

41) <https://greatfortune2u.tistory.com/28>

42) <https://www.etri.re.kr/webzine/20190215/sub01.html>

43) <https://www.etnews.com/20220919000157>

44) 양선진,박재표,양승민. (2019). 스마트 시티 환경에서 대규모 영상 모니터링을 위한 IP 비디오 월 시스템의 설계 및 구현. 『한국산학기술학회』, 20(9), 7-13.

45) <https://www.aitimes.kr/news/articleView.html?idxno=25562>

46) <https://www.cctvnews.co.kr/news/articleView.html?idxno=215145>



[그림1-7] 인공지능을 활용한 디지털 통합관제센터

친화적인 제어하는 방식으로 전환이 될 것으로 판단된다. 통합관제센터는 디지털화 발전하여 제조업은 물론 항공, 교통, 도시, 의학, 헬스 등 다양한 산업에 적용되고, AI, VR/AR, IoT, 빅데이터, 클라우드 등과도 뗄 수 없는 기존 기술의 바탕으로 만들어지고 있다.⁴⁷⁾

ETRI가 만든 기술 발전 지도로 2035년에 표현되는 미래의 디지털 트윈 사회의 자율형 도시 모습을 재현하는데,⁴⁸⁾ 현실과 가상이 공존하는 사회의 발전으로 기존의 제어방식인 마우스나 키보드의 방식이 점점 사라지고 멀티모달로 모니터링 및 제어되는 솔루션은 더욱 필요할 것으로 전망된다.

47) <https://www.kidd.co.kr/news/216503>

48) <https://www.aitimes.kr/news/articleView.html?idxno=25795>

제 3 절 연구의 범위와 목적

1) 연구 범위

가) 멀티모달 및 비디오월 정의

멀티모달(Multi-modality)은 음성이나, 제스처, 키보드 및 마우스와 같은 다양한 방식으로 인간과 컴퓨터가 2개 이상의 입출력 모달리티를 이용하여 상호 작용하는 환경이며, 의사소통 채널이다.⁴⁹⁾

멀티모달 인터페이스는 인간과 컴퓨터 상호 작용할 때 보다 인간 친화적인 인터페이스라고 할 수 있으며, 특정한 인터페이스에 종속되지 않는 다양한 입출력 방식으로 자연스러운 입출력 인터페이스의 결과를 얻을 수 있다.⁵⁰⁾

비디오월(Video Wall) 시스템은 다수의 상황판이나 대형 디스플레이 장치에 여러 영상을 동시에 자유롭게 재생하거나 표시할 수 있는 솔루션으로 스마트 시티의 통합관제센터의 다양한 영상을 제어하고 표출하는 솔루션으로 빌딩 관리, 아파트 단지 관리, 데이터 센터, 교통관제 시스템 등에 많이 사용된다. ⁵¹⁾

입출력의 상호작용을 하기 위한 인터페이스 방식은 멀티모달을 활용한 음성과 제스처, 음성과 눈동자 인식 등 2개 이상의 인간-컴퓨터 사이의 상호작용이 사용자와 시스템 사이의 인터페이스가 존재하여 다양한 미디어의 디스플레이인 컴퓨터뿐만 아니라 모바일, 셋톱박스, 미디어 스킨, 디지털 배너, 미디어 폴, 비디오월, 키오스크, 메뉴 보드 등 인터페이스의 다양성으로 확대되고 있다.⁵²⁾

기존의 키보드나 마우스 클릭의 전통적인 인터페이스 방식에서 멀티모달 방식인 손 모양과 음성 인식을 활용한 비디오월 제어시스템 연구이다.

49) Laurence Nigay, Joelle Cutaz. (1993). A design space for multimodal systems: concurrent processing data fusion, 『In Proc, INTERCHI'93, ACM Press』, 172.

50) 장복만, 김정석, 장효경, 박성도, 최의인.(2009). 멀티모달 인터페이스를 이용한 관제시스템. 『한국정보기술학회』, (6),934-938.

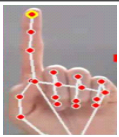
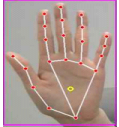
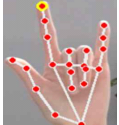
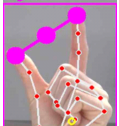
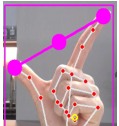
51) 양선진, 박재표, 양승민.(2019). 스마트 시티 환경에서 대규모 영상 모니터링을 위한 IP 비디오월 시스템의 설계 및 구현. 『한국산학기술학회』, 20(9),7-13.

52) 정유경.(2011). 멀티 디스플레이 사례분석과 관리 인터페이스에 관한 연구. 『디자인지식저널』,18,180-189.

2) 연구 모델 기능 도출

손 모양과 음성 인식의 멀티모달 인터페이스를 활용한 비디오월 시스템은 손 모양과 음성 인식 부분으로 비디오월 제어시스템의 UI 제어하는 시스템으로 나누어 기능을 도출하였다. [표1-2]와 같이 마우스 이벤트에 따라 손동작을 인식하여 컴퓨터 모니터 화면에 인식되어 음성과 동시에 화면을 제어하는 기능으로 정의하여 구성하였다.

[표 1-2] 연구 모델 상세 기능

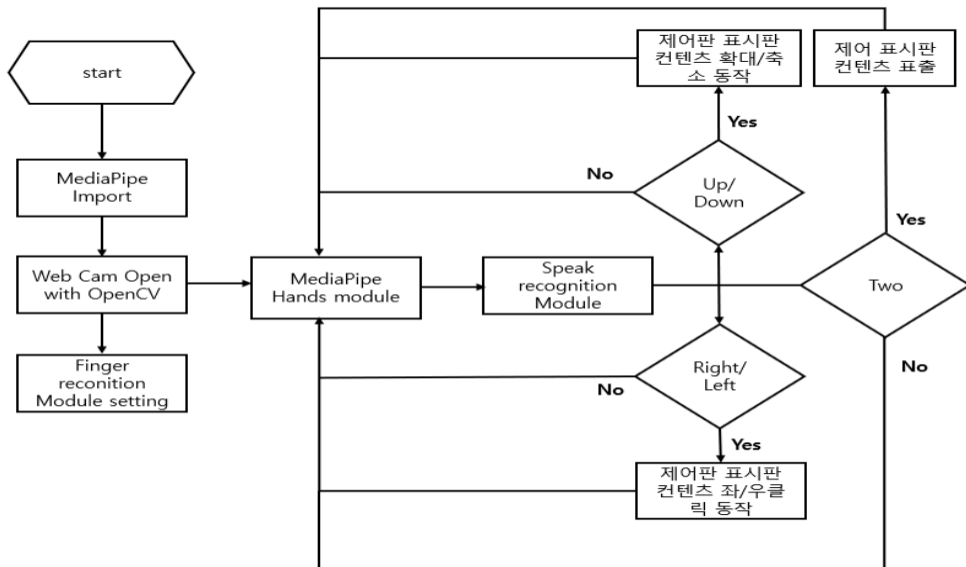
구분	상세 기능	
손동작 인식		- 모니터 화면에 검지 모양을 인식하여 마우스 포인트가 화면에 상하좌우로 이동 기능
		- 모니터 화면에 검지로 마우스 포인트를 이동한 후 손바닥을 펴면 더블 클릭 기능
		- 엄지를 펴고, 쥐는 손동작으로 마우스 좌/우 클릭 기능
		- 모니터 화면의 영상을 검지 모양 지정 후 검지와 엄지를 붙이면 영상이나 이미지 축소 기능
		- 모니터 화면의 영상을 검지를 지정한 후 검지와 엄지를 열면 이미지 확대 기능
음성 인식	- 검지로 콘텐츠를 지정한 후 음성명령으로 스크롤 바 업/다운 기능 - 검지로 아이콘에 지정된 상태에서 음성명령으로 'two' 하면 더블클릭 기능 - 검지로 이미지, 영상을 지정, 음성으로 'up', 'down' 음성명령을 내리면 콘텐츠가 축소, 확대 기능	
비디오월 제어시스템 UI 제어	- 검지로 비디오월 제어시스템 UI에 지정되어 음성명령 'two', 'right', 'left', 'up', 'down'으로 더블 클릭, 마우스 좌/우클릭, '콘텐츠 확대, 축소 가능 기능	

3) 연구 모델 시스템 흐름도 및 기술 요소

가) 연구 모델 시스템 흐름도

멀티모달 인터페이스를 이용한 비디오월 제어시스템은 음성 및 손 모양을 인식하여 사용자 인터페이스를 이용하여 제어하는 시스템이다. 본 시스템은 기존의 키보드나 마우스의 전통적인 방식의 인터페이스에서 손 모양과 음성을 인식하여 비디오월 시스템을 제어하는 시스템으로 손 모양의 5가지 부분을 정의하고 손 모양과 음성을 결합한 멀티모달 인터페이스를 통해 시스템이 UI를 제어하도록 시스템을 구성하였다.

[그림 1-8]은 본 시스템의 연구모델 시스템 흐름도이며, 웹 카메라로 영상을 캡처하여 손 모양을 MediaPipe로 인식 및 보정을 통해 모니터에 좌표를 획득하여 손 모양으로 마우스 포인트 이동 후 음성명령을 인식하여 인터페이스 되고, 기존 마우스 이동, 더블클릭, 드래그, 이미지 및 영상의 확대와 축소를 모양이나 음성으로 제어하는 시스템으로 구분하여 기본 연구모델을 설계하였다.



[그림 1-8] 연구 모델 시스템 흐름도

나) 연구 모델 기술 요소

멀티모달 인터페이스를 이용한 비디오월 제어시스템의 음성 및 손모양 인식에 대한 기능별 기술 요소는 [표 1-3]과 같다.

손 모양 인식에서 제스처 인식 및 모션 인식 기술이 적용되며 정의된 손 모양을 웹캠 기기를 통해 인공지능 인식 모델을 활용하여 비디오월 제어시스템의 UI와 인터페이스 되도록 설계하였고, 손 모양을 인식한 후 비디오월 제어시스템의 아이콘을 클릭하여 음성명령을 내려 콘텐츠가 동작하도록 구성하였다.

비디오월 제어시스템의 UI에서는 손모양 및 음성 인식으로 아이콘을 클릭하거나 리스트를 드래그하여 아이콘을 선택하여 기능 구현할 수 있도록 UI 아이콘을 지정하여 사용자가 최대한 편리하게 사용할 수 있도록 UI 아이콘을 분류하여 적용하였다. 디스플레이에 출력시 제어화면이 이미지 및 영상이 손 모양이나 음성으로 화면을 축소, 확대 등의 기능이 인터페이스가 되도록 구현되었다.

비디오월 제어시스템의 UI 기능 중에 음성 및 손 모양으로 제어가 가능한 아이콘이 사용자의 빈도수 등을 고려하여 제어하는 부분을 구분하였고 이미지나 영상의 소스 부분의 리스트에서 특정 소스를 불러와서 표출시 음성으로 명령하여 콘텐츠가 표출이 되도록 구성하였다.

[표 1-3] 연구모델 기능별 기술 요소

구현 항목	기술 요소	설명
손모양 인식	· 제스처 인식 기술 · 모션 인식 기술	· 웹캠으로 손모양 인식 · 정의된 손모양 인식 획득
음성 인식	· 자연어 처리 · 음성인식 인터페이스	· 음성 인식 인터페이스 구성 · 손모양 인식과 인터페이스 구성 · UI와 인터페이스 구성
비디오월 제어시스템 UI	· 음성 인식 인터페이스 · 손모양 인식 인터페이스	· UI 인터페이스 구성 · 음성 및 손모양 인식 인터페이스 구성

4) 연구의 목적

본 연구는 음성 및 손모양 인식을 이용한 멀티모달 인터페이스 방식의 비디오 월 제어하는 시스템으로 스마트 시티의 통합관제센터, 디지털 트윈을 활용한 디지털 통합관제센터에 적용하여 인간 친화적인 인터페이스를 통해 컴퓨터와 인간이 상호작용을 통해 편리하게 정보를 교환하는 것을 목적으로 한다.

다양한 IT 기기들이 인간과 상호작용할 때 한가지 방식의 인터페이스에서 두가지 이상의 입출력으로 인터랙션이 가능하게 발전하고 있다. 스마트 홈에서 가전을 제어하는 방식이 음성이나 제스처로 간단하게 제어하는 기술 등은 오래전부터 기술이 개발되어 상용화되고 있지만 보다 고도화되는 비디오월 제어시스템에는 아직 연구해야 할 부분이 많다.

손 모양 및 음성 인식을 활용하여 통합관제센터의 디스플레이를 제어하는 비디오월 제어시스템은 통합관제의 운영자가 더욱 편리하고 신속하게 제어할 수 있는 기술을 제공한다. 그리고 공공, 제조, 건설, 발전소 등 다양한 분야에서 활용이 가능하고 기능을 더욱 고도화할 필요가 있다.

LED 기술의 발전으로 통합관제센터의 구축되는 대형 디스플레이는 DID 모니터에서 1mm 이하 픽셀의 LED 디스플레이로 전환이 점점 가속화되고 있고, 최근 방송사에서 진행하는 XR 스튜디오를 이용한 가상 영상을 활용한 현지 촬영과 같은 효과를 XR 스튜디오에도 구현하고 있다.

이러한 형태의 LED 디스플레이의 활용은 멀티모달 인터페이스를 활용한 대화형 인터랙션 기술의 적용이 확대되고 있고, 멀티모달 인터페이스를 이용한 비디오월 제어시스템 기술이 인간 친화적이고 운영자가 운영환경을 개선하여 스마트 시티 및 다양한 산업에 영향을 미칠 수 있다.

제 2 장 이론적 배경 및 선행연구

제 1 절 손동작 인식을 활용한 상호작용

1) 손동작 인식에 대한 개요

사람과 컴퓨터의 인터페이스가 기존의 키보드, 마우스의 전통적인 방식에서 음성, 제스처 등 다양한 방법으로 인터랙션이 되고 있다. 여기서 제스처를 이용하는 방식에서 손동작은 일상적인 외부와 상호작용하는 가장 원초적인 도구로써, 손의 위치, 자세, 동작을 동시에 이해하여 일상생활에서 편의성을 증대시키는 중요한 인터페이스 방법이다.⁵³⁾

이러한 방식이 음성으로 전달하는 방식보다 비언어적인 수단을 이용하여 상대방에게 자기의 의사를 표현하는데 다양한 형태의 손짓, 몸동작, 얼굴 표정 등의 비언어적인 수단으로 전달 방식이 더 많이 활용되고 주변에 소음으로 인해 음성의 전달할 수 없는 공간에서는 효율적으로 전달할 수 있는 방식이다.⁵⁴⁾

컴퓨터와 인터페이스에서 음성 인식과 동시에 손 모양 및 동작을 인식하기 위한 다양한 기술이 개발되고, 다양한 손 모양에 관한 영상에서 추출된 특징 벡터 등의 데이터베이스를 만들어 두고 있다.⁵⁵⁾ 단순히 2D 패턴 매칭이나 특징 벡터 비교를 통해 손바닥 포즈 및 손 모양을 인식하기도 한다.⁵⁶⁾

손의 기하학적인 특성을 기반으로 손가락을 검출하여, 손가락의 수로부터 손 모양을 인식하여 손의 정확한 3D 모델을 미리 가지고 있다고 가정

53) 김동욱, 이선경, 정찬양, 이창화, & 백승렬. (2020). 손 위치, 자세, 동작의 통합 심층 학습. 『한국정보처리학회』, 27(2), 1048-1051.

54) J.Ohya and Y.Kitamura, etc.(1993). Real-Time Reproduction of 3D Human Images in Virtual Space Teleconferencing, in Proc. Virtual Reality Annual Int. Symp. 408-414,

55) Cao, C., Sun, Y., Li, R., & Chen, L. (2011). Hand posture recognition via joint feature sparse representation. Optical Engineering, 50(12), 127210-127210.

56) Choi, J., Park, H., & Park, J. I. (2011). Hand shape recognition using distance transform and shape decomposition. International Conference on Image Processing (pp. 3605-3608). IEEE.

하고, 손 모델을 카메라 영상 상의 2D 손 영상과 매칭시키는 방법으로 손 모양을 인식하는 것으로 연구되었다.⁵⁷⁾⁵⁸⁾

이러한 연구는 손의 위치, 자세, 동작을 이해하는 연구가 상호 작용에 대한 고려 없이 정답이 주어졌다고 가정한 채 실험실 환경의 연구 진행이 되는 초기 발전의 단계에 있다.

반면에, 응용적인 면에서 마이크로소프트사의 홀로렌즈(Hololens), 페이스북이 인수한 오쿨러스사의 리프트(Rift), 구글 글래스, 삼성 기어VR 등 다수의 스마트 글래스에서 채택되는 1인칭 시점에서 손의 위치, 자세, 동작의 이해는 중요한 문제로 인식되어 발전하고 있다.⁵⁹⁾

또한 손동작 인식을 이용한 스마트 디바이스들이 2006년 닌텐도 Wii를 시작으로 2010년에 출시된 동작 인식을 새롭게 한 마이크로소프트사의 Kinect는 맨손으로 사용이 가능하여 영화 마이너리 리포트보다 훨씬 편리한 사용자 경험을 제공하였다.

동작 인식 기능이 리모콘에 탑재되어 제어되는 스마트 TV, 손의 동작을 감지하는 스마트 반지, 근육의 변화를 감지하는 제스처 인식 디바이스, 2개의 카메라를 사용하는 립모션 제품 등이 응용되어 상용화되고 있다.⁶⁰⁾ 향후의 IT 제품들은 사용자 경험을 극대화하기 위해 자연스럽고, 직관적이며, 즉시성을 더욱 강조할 것으로 전망된다.

2) 손동작 인식에 대한 기존 연구

손동작 인식 기술은 컴퓨터가 사람의 손가락이나 손동작을 인식하기 위해서는 손가락의 길이와 위치, 손바닥 및 팔의 각도 등 사람이 무의식적으로

57) Choi, J., Park, H., & Park, J. I. (2011). Hand shape recognition using distance transform and shape decomposition. International Conference on Image Processing (pp. 3605-3608). IEEE.

58) 박한훈, 최준영, 박종일, & 문광식. (2013). Kinect 기반 손 모양 인식을 위한 손 영역 검출에 관한 연구. 『방송공학회논문지』, 18(3), 393-400.

59) 김동욱, 이선경, 정찬양, 이창화, & 백승렬. (2020). 손 위치, 자세, 동작의 통합 심층 학습. 『한국정보처리학회』, 27(2), 1048-1051.

60) <https://www.epnc.co.kr/news/articleView.html?idxno=13427>

인식되는 것처럼 컴퓨터도 일련의 프로세스를 거쳐 모듈이 실행된다.

또한 사용자가 미리 원하는 명령을 설정하여 손의 상태에 따라 카메라 영상을 추출하여 미리 설정된 명령어를 프로그램에 저장하여 일정 프레임을 학습하여 손 상태를 확인하여 명령을 수행한다.⁶¹⁾

사람의 손동작 중 검지만을 펴서 표현하는 포인팅 동작은 방향을 지시하기 위하여 가장 많이 사용되는 동작이며, 손동작을 추적하여 윈도우 시스템을 위한 마우스 포인터를 제어할 수 있다면 사용자는 자연스럽게 컴퓨터에 의사 전달이 가능하다.⁶²⁾ 컴퓨터의 모니터 해상도가 커지면서 웹 카메라와 모니터 간의 해상도 차이로 단순한 웹 카메라에서 손의 위치를 인식하여 포인터를 이동시키는 방법으로는 모든 범위에 커서를 위치시킬 수 없다.

따라서 되돌아오는 동작을 반복하지 않고, 마우스의 위치를 모니터의 가장 자리까지 원하는 곳으로 편리하게 위치시키고, 엄지 및 검지 손가락의 움직임으로 윈도우 시스템의 마우스를 제어하는 방식으로 인식률이 97% 이상의 유효성을 확인하였다.⁶³⁾

선행 학습이 필요 없이 손동작 인터페이스 구현하는 방법으로 1대의 웹캠과 오픈소스 영상처리 라이브러리 OpenCv를 사용하여 차 영상과 화소값 기반의 영상처리 과정을 통해 실시간으로 손영역을 추적, 이진화시켜 손가락의 움직임도 값이 변화지 않도록 중심 모멘트를 설정하여 마우스 커서 움직임을 활용하는 방법으로 핸드마우스 인식률은 86.5%에서 94%의 결과를 얻었다.⁶⁴⁾

또한 손모양을 인식하는 방법 중 다수의 카메라를 이용한 픽셀 시차 값을 계산하는 방식은 많은 연산량 때문에 실시간 응용에 한계가 있고, 컬러 영상 카메라를 이용하는 방식도 배경 및 조명변화에 민감하여 인식률이 제한적이다. 깊이 영상 카메라를 이용하여 배경 조명의 변화에 강인한 손검출 방법을

61) 김희성, 이봉환, & 이규원. (2014). 손동작 인식을 통한 컴퓨터 제어. 『한국정보처리학회』, 21(1), 912-915.

62) 하유솔, 고은지, 김명준.(2011).손동작 인식 기반의 마우스 입력 장치. 『한국정보처리학회』, 38(10),524-535.

63) 김정인. (2012). 손동작 인식을 이용한 마우스제어기법. 『멀티미디어학회』, 15(11), 1377-1383.

64) 강성원, 김철중, & 손원. (2009). 손 동작 인식을 통한 사용자에게 편리한 핸드마우스 인터페이스 구현. 『한국방송미디어공학회』, 129-132.

이용하여 컴퓨터의 마우스를 제어하는 연구 방법으로 실시간 손영역의 중심점을 효과적으로 획득하고, 손가락 끝점을 찾는 방법을 구현하였다.⁶⁵⁾

이처럼 손동작 인식 기술은 다양하게 제안되었고 손동작 인식으로 잘 알려진 Viola Jones 방법은 파란색 직사각형으로 손영역을 찾아 정확하게 인지할 수 있고, 특징 추출을 위해 Haar와 유사한 특징 세트를 이미지 표현에 통합할 수 있다.⁶⁶⁾

그리고 비전 기반 손 추적 접근 방식을 사용하여 손동작 분할에 사용되었고,⁶⁷⁾ SOM(Self-Organizing Map)은 5DT 데이터 장갑을 사용한 손동작의 또 다른 기술로 데이터를 입력 벡터로 사용한 다음 데이터와 동작 명령간의 매핑하는 기계 학습 알고리즘의 형태로 연구가 진행되었다.⁶⁸⁾

또한 색상과 손짓 깊이 이미지를 추출하는 기술은 손 영역은 컬러필터를 사용하여 필터링하고, 손바닥에 작은 센서를 기반으로 검색공간을 줄이고 강력하고 계산하기 쉬운 인식 방법으로 3D 각도 포즈를 제공하는 미세한 가속도계와 자력계를 손바닥에 작은 센서를 부착하여 이미지 포인터, 깊이, 손방향, 피부색 등을 고려한 연구도 진행되었다.⁶⁹⁾

손동작을 이용하여 소리와 음악을 제어하는 연구에서는 TDNN(Time-delay Neural Network)통해 인식되는 손 제스처를 통해 음악 및 사운드 매개변수를 제어하는 접근 방식을 제시하여 음악 제어를 위해 왼손을 사용하여 왼쪽 가리키기, 오른쪽 가리키기, 악수, 위아래로 움직이는

65) 박세호, 김태근, 이지은, & 이정택. (2014). 손 제스처 인식을 위한 손 영역 추적 및 손가락 검출 방법. 『한국방송미디어공학회』, 34-35.

66) Yun, L., & Peng, Z. (2009). An automatic hand gesture recognition system based on Viola-Jones method and SVMs. In 2009 Second international workshop on computer science and engineering (Vol. 2, pp. 72-76).

67) Xu, H., Hou, X., Su, R., & Ni, Q. (2009). Real-time hand gesture recognition system based on associative processors. In 2009 2nd IEEE International Conference on Computer Science and Information Technology (pp. 14-18).

68) Jin, S., Li, Y., Lu, G. M., Luo, J. X., Chen, W. D., & Zheng, X. X. (2011). SOM-based hand gesture recognition for virtual interactions. In 2011 IEEE International Symposium on VR Innovation (pp. 317-322).

69) Trindade, P., Lobo, J., & Barreto, J. P. (2012). Hand gesture recognition using color and depth images enhanced with hand angular pose data. In 2012 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI) (pp. 71-76). IEEE.

등 다양한 제스처를 구별하기도 하였다.⁷⁰⁾

또한 모바일 장치에서 손의 인식을 개선하기 위해 HOG(Histogram of Oriented Gradients)와 LBP(Local Binary Pattern)를 결합하여 인식률 92% 정확도를 달성하기도 했다.⁷¹⁾ 손동작을 인식하여 컴퓨터를 제어하는 연구는 손동작의 위치나 인터페이스 기기에 따라 약간의 차이는 있지만 인식률은 90% 이상이며 적외선 카메라를 이용하여 검지를 편 상태에서 움직이면 검지 끝을 따라 포인터를 이동하는 인식률은 100%의 인식률이 나오는 연구도 진행되고 있다.⁷²⁾

다양한 손동작이 다양한 기기를 통해 컴퓨터가 제어되는데 일부 동작은 인식률은 100%로 인식하고 있지만 대부분 동작이 90%대에서 인식 및 제어가 되는 것을 볼 수 있다. 추가적인 연구를 통해 더 많은 손동작이 컴퓨터를 제어할 때 인식률을 높이는 연구의 필요성이 대두되고 있다.

3) 컴퓨터 제어를 위한 손동작 형태

손동작은 가장 큰 범주에서 여러 가지 커뮤니케이션하는 수단으로 많이 사용하고 있다.⁷³⁾ 손모양을 분류하여 다양한 애플리케이션 시나리오에 대화 제스처, 제어, 조작 및 의사소통 제스처로 나눌 수 있고, 컴퓨터 상호작용에 대한 부분은 제어 제스처에 속하는 것이 적합하다.⁷⁴⁾

손동작만으로 컴퓨터상의 응용 프로그램을 제어할 수 있는 휴먼 마우스 연구에서 마우스 포인트 이동은 손을 편 상태, 마우스 좌클릭은 손을 편 상태에

70) Modler, P., & Myatt, T. (2008). Recognition of separate hand gestures by time-delay neural networks based on multi-state spectral image patterns from cyclic hand movements. In 2008 IEEE International Conference on Systems, Man and Cybernetics (pp. 1539-1544).

71) Lahiani, H., & Neji, M. (2018). Hand gesture recognition method based on HOG-LBP features for mobile devices. *Procedia Computer Science*, 126, 254-263.

72) 김효성, 김병규, 정현기.(2011). 적외선 카메라를 이용한 에어인터페이스 시스템(AIS) 연구. 『정보처리학회지』, 18(3), 109-116.

73) Gesture Internet Document (2007) From “<http://www.en.Wikipedia.org/wiki/Gesture>”

74) Wu Y, Huang TS (1999) Vision-based gesture recognition: a review. International Gesture Workshop, France.

서 네 개가 손가락을 접는 모양으로 클릭 기능을 구현한 연구이다.

마우스 드래그는 손을 편 상태에서 엄지와 검지만 남기고 손가락을 펴면 마우스 드래그가 되고, 마우스 우클릭은 손을 편 상태에서 약지와 새끼손가락을 접으면 우클릭이 되며, 마우스다운/엮은 엄지손가락을 검지를 펴면 다운, 검지와 중지를 펴면 엮이 되는 기능을 구현하는 손동작 연구도 있다.⁷⁵⁾

2차원 입력 영상으로부터 동적인 손동작 없이 손 특징을 이용한 손동작 13가지에 인식에 연구도 사람에 따라 손의 종류 및 동작 난이도 차이를 보여주고 있다.⁷⁶⁾ 검지 하나만으로 마우스 포인트 이동, 좌클릭, 우클릭, 더블클릭의 마우스 이벤트를 수행하는 구현에서 인식률이 평균 90% 이상 인식된다.⁷⁷⁾ 다양한 마우스 이벤트에서 엄지와 검지를 이용한 손가락 동작 인식에서 마우스 이동, 마우스 좌클릭, 마우스 우클릭, 더블클릭, 드래그를 수행하였다.⁷⁸⁾

수화를 기반으로 간단한 손동작으로 리모컨을 제어한 12개의 손가락 패턴으로 전원, 채널 변경 신호 전달이 가능함을 확인하였다.⁷⁹⁾ 원격에서 전자칠판을 제어하는 다양한 손 모양을 검출하고 추적하는 구현에서 마우스 이동 손바닥 편 모양을, 마우스 더블클릭은 주먹의 손 모양을, 마우스 드래그는 엄지와 검지를 편 가위 모양으로 효과적인 제어가 가능하다.⁸⁰⁾

증강현실 응용을 위한 마우스 버튼 클릭의 역할을 손바닥 무게중심에 대한 엄지와 검지의 끝점 각도를 계산하여 엄지와 검지의 떨어짐으로 동작을 인식하는 기법도 있다.⁸¹⁾ 모바일 기기의 카메라 장치에서 실시간 영상을 통해 실

75) 정지인, 김혜림, 서세현, 계이기, 최황규. (2009). 손동작 인식을 이용한 프로그램 제어 인터페이스 설계 및 구현. 『한국정보처리학회지』, 16(2), 383-384

76) 이래경, 김성신. (2001). 손동작 인식을 통한 Human-Computer 구현. 『한국지능시스템학회』 11(1), 28-32.

77) 강성원, 김철중, 손원. (2009). 손 동작 인식을 통한 사용자에게 편리한 핸드마우스 인터페이스 구현. 『한국방송미디어공학회』, 129-132.

78) 김정인. (2012). 손동작 인식을 이용한 마우스제어기법. 『멀티미디어학회』, 15(11), 1377-1383.

79) Choi, C. S., & Oh, K. S. (2010). Gesture Recognition Algorithm from the Webcam Images. In *Proceedings of the KAIS Fall Conference* (pp. 88-91). The Korea Academia-Industrial cooperation Society.

80) 박상윤, 이응주. (2010). 복잡한 영상에 강인한 손동작 인식 방법. 『멀티미디어학회지』, 13(7), 1000-1015.

세계에서 상호 작용하는 방식에서 다양한 손모양으로 tip 조작, collision mode의 select 조작, finger over mode의 select 조작, grasp의 조작, spread 조작, release 조작을 구현하였다.⁸²⁾

그리고 별도의 부착물 없이 키넥트 카메라만을 사용하여 퍼진 손가락 수에 따라 0개는 마우스 누름, 1-2개는 마우스 움직임, 4-5개는 마우스 올림으로 손동작을 인식하여 다음페이지, 이전페이지, 아래로, 위로 커서 움직임 및 클릭을 동작하였다.⁸³⁾

미디어파이프 핸드 모델을 이용하여 손 포인트, 손 마우스, 손 다이얼 및 손필기 제스처를 인식하여 스마 홈 기기를 제어하고,⁸⁴⁾ 딥러닝을 기반하여 실제 마우스 없이 좌클릭, 우클릭, 스크롤 기능으로 컴퓨터를 제어하는 연구도 진행되었다.⁸⁵⁾

다양한 형태의 손동작으로 마우스의 기능을 구현하여 컴퓨터를 제어한 연구들이 꾸준히 진행되고 있지만 가장 적합한 손동작으로 제어하는 기준이 없다는 것을 알 수 있다. [그림 2-1]이 같이 손동작과 이벤트를 구분하여 연구한 것으로 마우스 이동이나 커서 이동은 검지로 활용한 부분이 많고, 드래그, 확대 축소 등의 이벤트는 두 손가락을 이용하기도 하고 양손으로 동작을 활용하여 연구한 사례이다.

손동작은 동작이 어려운 동작은 오히려 컴퓨터를 제어하는 데 부담이 되는 부분이 있으므로 가능한 쉬운 동작을 기준으로 반영이 하는 것이 효과적이다. 일상생활에서 물건을 지정하거나 컴퓨터 제어나 TV 리모콘으로 on/off할 때 주로 사용하는 손가락이 검지인 것을 확인할 수 있다.

81) 이정진,김종호,김태영. (2010). 증강현실 응용을 위한 손끝점 추출과 손동작 인식 기법. 『멀티미디어학회지』, 13(2), 316-323.

82) Patel, N., & He, S. J. (2018). A survey on hand gesture recognition techniques, methods and tools. Int J Res Advent Technol, 6(6).

83) Sang-Hun Lee, Dae-jin Kim, & Hong-Sub Choi (2012). Implementation of DID interface using gesture recognition. Journal of Digital Contents Society, 13(3), 343-352

84) 송락빈, 홍윤아, 곽노윤. (2023). MediaPipe Hands 모델 기반의 손 제스처 인식을 이용한 사용자 인터페이스. 『멀티미디어학회논문지』, 26(2), 103-115.

85) Shriram, S., Nagaraj, B., Jaya, J., Shankar, S., & Ajay, P. (2021). Deep learning-based real-time AI virtual mouse system using computer vision to avoid COVID-19 spread. Journal of healthcare engineering, 2021

정지인, 김혜림, 서세현, 제이기, 최황규. (2009).	동작						
	이벤트	마우스 포인트 이동	마우스 좌 클릭	마우스 드래그	마우스 우 클릭	스크롤 마운	스크롤 업
김효성, 김병규, 정현기. (2011)	동작						
	이벤트	마우스 포인트 이동	마우스 드래그	마우스 클릭	마우스 휠 업(확대)	마우스 휠 다운(축소)	
강성원, 김철중, 손 원(2009)	동작						
	이벤트	마우스 포인트 이동	마우스 좌 클릭	마우스 우 클릭	마우스 더블 클릭		
김정인(2012)	동작						
	이벤트	마우스 이동	마우스 좌클릭	마우스 우클릭	마우스 더블 클릭	마우스 드래그	
이세인(2008)	동작						
	이벤트	출동발생	객체 선택	객체 해방	움켜 잡기	모드변경	모드변경
Shriram, S., Nagaraj, B., Jaya, J., Shankar, S., & Ajay, P. (2021).	동작						
	이벤트	커서 이동 및 멈춤	손가락 감지	마우스 이동	마우스 좌 클릭	마우스 우클릭	스크롤 다운

[그림 2-1] 손동작 및 마우스 이벤트

4) 손동작 인식으로 산업 활용 사례

2010년 11월에 마이크로소프트가 x-box 게임기에 연결하여 사용할 수 있는 ‘키넥트’라는 동작센서 제품을 출시함으로써 장갑을 사용하지 않고 손동작으로 컴퓨터의 게임을 제어할 수 있고, 립 모션사의 제스처 기반 인간 인터페이스 장치도 다양한 손동작으로 향상된 사용자 경험을 제공하고 있다.⁸⁶⁾

[그림 2-2]와 같이 logbar사의 제품인 더링은 검지 손가락에 맞는 웨어러블 디바이스로 허공에 손가락으로 글씨나 아이콘 형태의 간단한 그림을 그리면 이를 인식하며, 제스처 제어 기능, 문자 전송, 결제 정보 전송, 알림 기능 등의 4가지 기능을 갖추고 있다.⁸⁷⁾

86) <https://www.epnc.co.kr/news/articleView.html?idxno=13427>

87) <https://vandrico.com/wearables/device/logbar-ring.html>



[그림 2-2] logbar사의 스마트 반지와 윈도우 기반 3D핸드 트래커

또한 화면에 손을 대지 않고 제스처로 간단한 기능을 조작하는 제스처 컨트롤을 이용하여 손짓만으로 차문을 열고 닫거나 차시트 각도를 조절하고, 선루프를 여는 단계까지 가능하여 조만간 K사에서 상용될 것으로 전망된다.

S사의 한 스마트폰 모델은 디지털 스타일러스 펜을 이용한 제스처로 스마트폰 기능을 제어하며 멀티미디어 기능을 극대화시켜 디지털 펜을 누른채 왼쪽에서 오른쪽으로 펜을 움직이자 휴대폰을 직접 터치하지 않고도 리모트 모드로 컨트롤이 가능하고 펜을 눌러 원을 그리면 좌우방향에 따라 줌인/아웃 가능하다.⁸⁸⁾

대형 멀티비디어 디스플레이에서 제스처를 위한 비디오 제스처 제어 기술 즉 손동작이나 몸을 움직이는 것만으로 멀티미디어 콘텐츠를 참여하고, 정보에 액세스하고, 특수효과를 조작하며 대화형 멀티비디어 디스플레이, 키오스크, 전시물, 디지털 간판 및 광고, 가상 게임시스템 및 대화형 표면 컴퓨팅 솔루션 등에 적용되고 있다.⁸⁹⁾

[그림 2-2]과 같이 카메라로부터 거리와 조정 가능한 볼륨 내에서 사용자의 손의 위치를 측정하여 윈도우 기반 애플리케이션에서 터치없이 커서를 제어 가능하고 추적기는 한 손이나 두 손을 따라가며 멀티터치 모드를 생성할 수 있으며 최대 10개의 XYZ 좌표를 동시에 추적할 수 있다.⁹⁰⁾

88) <https://blog.naver.com/mosfnet/221988404492>

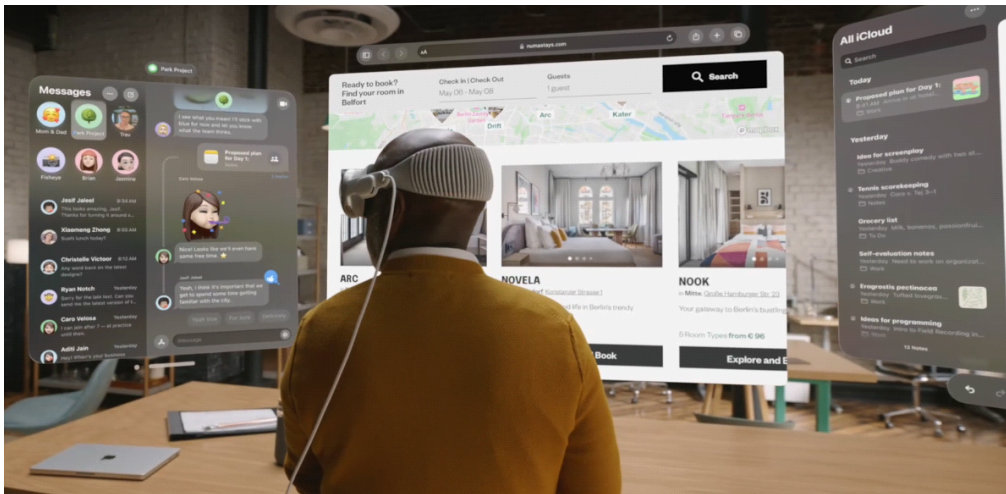
89) <https://www.gesturetek.com/>

90) <https://gesturetekhealth.com/products/gestrack-3d>

Motion Gesture만의 알고리즘을 이용한 손의 마디와 관절을 최소 6개에서 최대 12개의 포인트로 구분하여 모양과 움직임을 분석하여 다양한 종류의 카메라, 하드웨어를 지원하여 자동차, 항공, 리테일, 게임, 드론, 가전, 스마트 홈 등에 적용하여 활용도를 높이는 연구가 진행되고 있다.⁹¹⁾

최근 애플사에서 출시한 애플 비전프로는 최초의 공간 컴퓨터이다. 물리적인 공간을 넘어선 몰입감 만점의 환경 기능을 제공하고 손동작과 음성으로 콘텐츠를 제어할 수 3D 인터페이스 선보였다.⁹²⁾

이처럼 손동작 및 음성 인식을 이용한 솔루션이 다양한 분야에서 적용되고 활용되며, 관제센터의 제어 분야에서도 3D 비전 프로와 같은 솔루션으로 적용되어 더욱 몰입감이 높은 시스템으로 발전될 전망이다.



[그림 2-3] 애플사 비전 프로 동작 장면

91) <https://www.mdstech.co.kr/reference/4934>

92) <https://www.apple.com/kr/newsroom/2023/06/introducing-apple-vision-pro/>

제 2 절 음성 인식에 활용한 상호작용 관한 연구

1) 음성 인식 기술의 개요

음성 인식이란 사람이 말하는 음성 언어를 컴퓨터가 해석해 그 내용을 문자 데이터로 전환하는 처리를 말한다.⁹³⁾ 또한 전화, 마이크 등을 컴퓨터에 음성 인식 시스템으로 전달된 사람의 음성의 특징을 추출하고 분석하여 미리 입력된 인식 목록에서 가장 근접한 결과를 찾는 기술이다.⁹⁴⁾ 음성 인식 기술은 컴퓨터가 인간의 말을 알아듣는 기술로 50년 이상 꾸준히 연구되어 온 상용화된 기술로 지속적인 발전을 거듭하고 있다.⁹⁵⁾

음성 인식 기술은 사람의 목소리를 텍스트로 변환하거나 인식해 기계가 반응할 수 있는 기능으로 IBM에서 1962년 ‘Shoebbox’를 시작으로 음성 인식 기술이 발전에 중요한 역할을 해 왔으며, [표 2-1]과 같이 단계에 따라 1세대에서 4세대로 나눈다.⁹⁶⁾⁹⁷⁾

[표 2-1] 음성 인식 발전 과정

세대	연도	주요 역사
1세대	1952년	- 미국 AT&T Bell 연구소, 숫자 인식 시스템 Audrey 개발
	1962년	- IBM, 영단어 인식 및 숫자 계산이 가능한 Shoebbox 공개
2세대	1970년	- 국방첨단연구 사업국, 대규모 음성이해 연구 프로젝트 진행
3세대	1980년	- IBM, Hidden Markov Model을 이용한 대규모 음성시스템을 개발
3.5세대	1990년	- 최초 소비자용 Dragon Dictate 출시 - Bell South, ASR 서비스 시작 (수화기에서 말한 내용 인식) - Dragon Naturally Speaking 출시 (연속 음성인식 제품)
4세대	2007년	- Apple, Google을 선두로 개인비서 서비스 출시, 본격 상용화

93) https://ko.wikipedia.org/wiki/음성_인식

94) <https://www.callcenter.or.kr>

95) 이윤근. (2012). 음성인터페이스 기술 개요 및 스마트폰 환경에서의 서비스 동향. 한국통신학회지(정보와통신), 29(4), 3-9.

96) Pinola, M. (2017). History of voice recognition: from Audrey to Siri. itbusiness. ca.

97) 김건우, 윤재영. (2021). 인공지능 스피커의 디스플레이 탑재가 사용자 경험에 미치는 영향. 『Archives of Design Research』, 34(3), 173-190..

음성은 인간이 사용하는 가장 보편적이고 편리한 정보전달의 수단이며, 의사소통의 수단뿐만 아니라 음성을 이용한 기계나 장치를 동작하는 수단으로 중요한 역할을 하고 있다.⁹⁸⁾ 음성 인식은 AI와 머신러닝을 활용하여 문법, 구조, 오디오 신호 등을 이해하고 처리하는 수준까지 왔으며, 더욱 고도화되어 다양한 기기들을 음성 인식으로 제어하는 대화형 음성 인식 시스템으로 발전을 넘어 대중화 단계로 진입할 것이다.

2) 음성 인식에 대한 기술

음성 인식 기술은 음성 인식과 화자 인식 기술로 분류되고, 음성 인식은 다시 특정 화자에 대해서만 인식하는 화자종속 시스템과 화자에 상관없이 인식하는 화자 독립시스템으로 나누며, 화자종속 음성 인식은 사용 전에, 사용자의 음성을 저장, 등록시켜 실제 인식을 수행할 때 입력된 음성 패턴과 저장된 음성의 패턴을 비교하여 인식한다.⁹⁹⁾[표 2-2]

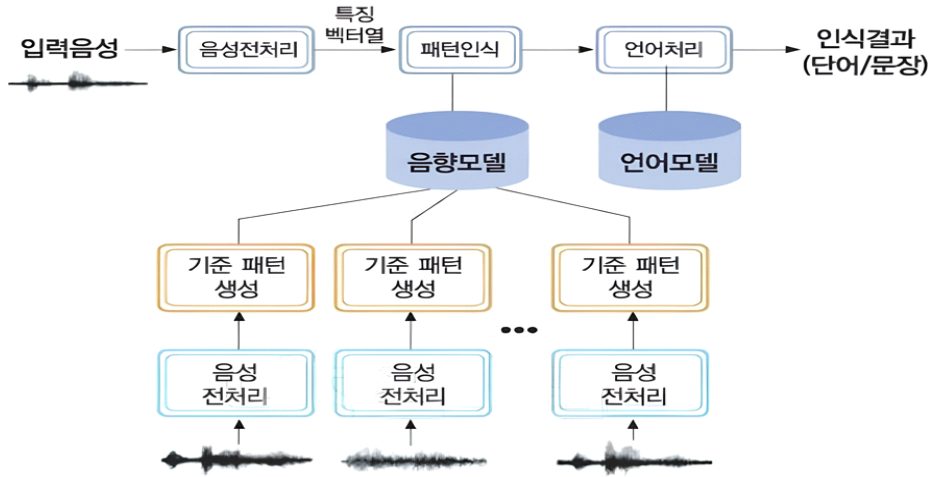
[표 2-2] 음성 인식 시스템 적용

분류	내용	적용	사례
적용 화자	화자종속	- 미리 등록한 특정화자만 인식 - 인식성능이 뛰어남	- 휴대폰 Voice dialing
	화자독립	- 불특정 화자를 대상 - 대용량의 음성 DB 필요	- Directory assistance 서비스
발음 형태	고립어	- 고립어 고립 단어 인식 - 단어 앞뒤에 묵음구간 존재	- Voice 브라우저
	연속어	- 연결단어, 연속문장 및 대화체 음성 인식 - 다양한 발음 변이를 기준 패턴과 Task에 맞는 언어모델 필요 - 핵심어 인식기술 요구(적용따라)	- 음성 타이프라이터 - 자동예약시스템 - 대화형 컴퓨터
어휘 수량	소용량	- 수백 단어 이하, 문맥 독립형 모델 사용	- pc commander
	대용량	- 수만 단어 이상, Sub-word 형태 모델 사용	- 음성 타이프라이터, 방송 뉴스 녹화

98) 노진우. (2009). 음성 인식을 위한 전처리 방법에 관한 연구. 『고려대학교 대학원』, 서울

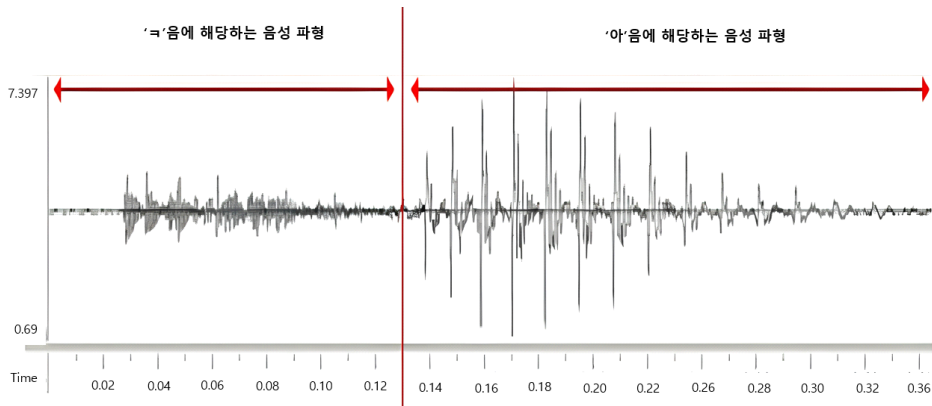
99) <https://wiki.wikisecurity.net/identification:음성인식>

음성 인식 기술의 원리는 [그림 2-4]와 같이 기본적으로 음성에 대한 파형을 분석하여, 1/1000초 단위로 작게 잘라 그 파형 신호 처리를 통해 10개 이상의 숫자로 변환하여 데이터의 특정 패턴을 인식하여 언어를 처리한다.¹⁰⁰⁾



[그림 2-4] 음성 인식 과정

[그림 2-5]는 카라는 발음에 대한 음성 파형이 앞에는 ㄴ 음이고 이어서 아 음이 연속되어 카 음이 되며, 음성 파형에서 주기적으로 나타나는 부분들이 하나의 패턴이 되어 소리로 인식하는 것이다.¹⁰¹⁾



[그림 2-5] 카 하는 글자를 발생시 음성 파형

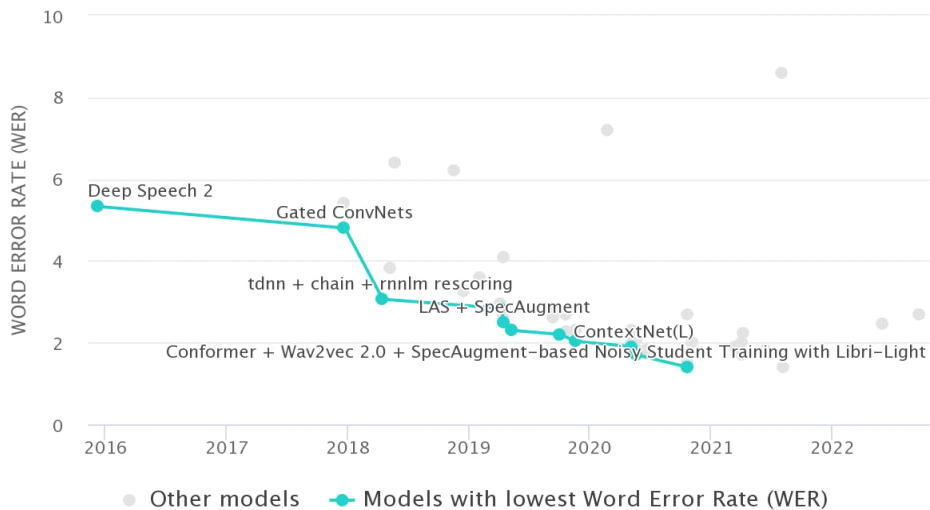
100) <http://computing.or.kr/14792/speech-recognition음성-인식/>

101) <https://blog.naver.com/msnayana/80199112243>

1980년대부터 음성 인식의 흐름은 템플릿(template) 방식의 접근 방식에서 은닉 마코프 모델 (hidden Markov model, HMM)과 같은 통계적 방법으로 기술의 변화를 가져왔으며, 전통적으로 음성 인식(Speech Recognition)에 가장 많이 사용되는 알고리즘이다.

이중 통계적 모델로서, 기본이 되는 음소열의 생성과 프레임 단위의 표면적 음향학적인 표현을 Markov 과정과 같이 확률로 나타냈고, 프레임 단위의 점수를 예측하는데 신경 회로망 (Neural network)이 사용되며, HMM 시스템과 결합되어 혼합 모델로 GMM-HMM(Gaussian Mixture Model-Hidden Markov Model)서 사용되었다.¹⁰²⁾

이 모델은 음성 인식과 같은 시퀀스 데이터 처리에 널리 사용되는 기계학습 모델로 HMM은 시퀀스 데이터의 시간적 변화를 모델링하여 특정 음성의 단위를 나타내는데 사용되고, GMM은 HMM 상태에 대응하는 음성 특성의 모델링한다.¹⁰³⁾¹⁰⁴⁾



[그림 2-6] 모델별 단어 수준 오류율

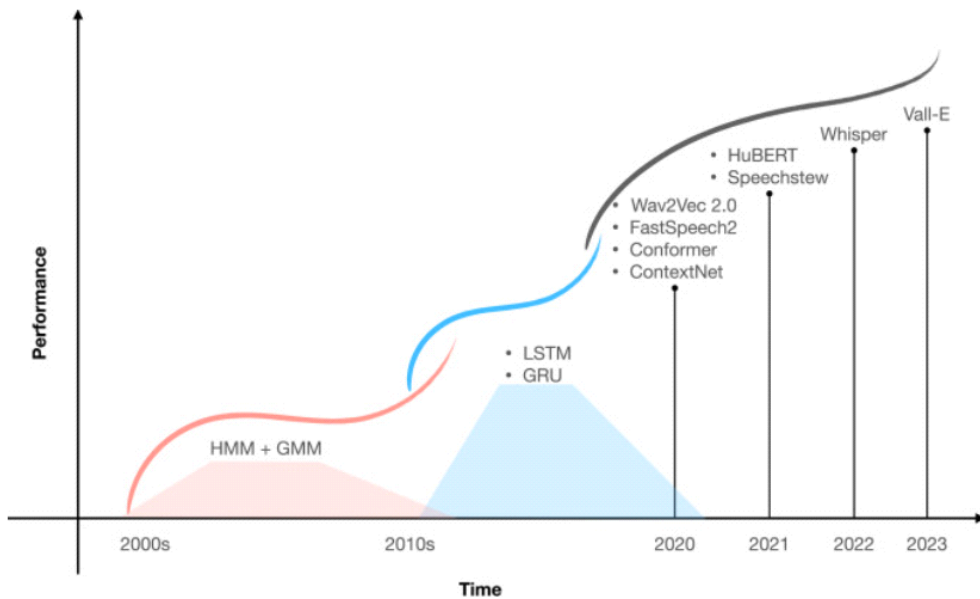
102) <https://blog.naver.com/msnayana/80102408151>

103) Rabiner, L. R., & Juang, B. H. (2006). Speech recognition: Statistical methods. Encyclopedia of language & linguistics, 1-18.

104) Huang, X., Acero, A., Hon, H. W., & Reddy, R. (2001). *Spoken language processing: A guide to theory, algorithm, and system development*. Prentice hall PTR.

초기에 음성 인식 시스템에서 이 모델이 효과적이었으나 최근에는 심층 신경망(Deep Neural Network, DNN) 모델로 대체되고 있다.¹⁰⁵⁾ 이 모델을 사용하는 딥러닝이 전환점을 작용하면서 음성 인식 또는 인공지능의 급격한 발전을 가져왔고, 기존의 표준화된 데이터에서 인간의 수준을 넘어선 성능으로, 산업과 일상에 적용하는 기술로 발전되었다.

[그림2-6]과 같이 WER(단어 수준의 오류)가 낮을수록 정확하다는 것을 보여 주며 Deep Speech 2 논문에서 인간은 5.8%였으나 4년 만에 인간에 비해 오류가 두 배의 적은 수치를 달성한 것을 볼 수 있다.¹⁰⁶⁾



[그림 2-7] 연도별 음성 인식 모델

그리고 [그림2-7]과 같이 MFCC(Mel-frequency cepstral coefficient) 및 HMM(Hidden Markov Models)과 같은 초기 접근에서 심층 신경망(DNN), 컨볼루션 신경망(CNN), 순환 신경망(RNN), 트랜스포머(Transformer), 컨포

105) Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., ... & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal processing magazine*, 29(6), 82-97.

106) <https://paperswithcode.com/sota/speech-recognition-on-librispeech-test-clean>

며(Conformer), 확산 모델(Diffusion model) 등으로 해마다 음성인식 모델의 진화과정을 보여주고 있다.¹⁰⁷⁾ 이러한 딥러닝 아키텍처의 급속한 발전은 음성 인식 기술의 분석, 제어 및 처리의 혁명을 가져왔다.

3) 음성 인식 기술의 산업 활용 사례

음성 인식 기술은 스마트홈, 의료산업, 자동 산업, 고객 서비스 분야 등에 다양한 분야에 활용되고 있다. [표2-3]는 음성 인식 기술의 활용 사례이다. 음성 인식 기술로 스마트홈에서 사용자는 음성 명령을 통해 스마트홈 기기를 제어할 수 있는데, 예를 들어, 스마트 조명을 켜거나 끄고, 온도를 조절하고, 스마트 TV를 제어하는 것이 가능합니다.

아마존 에코(Amazon Echo)나 구글 홈(Google Home) 같은 스마트 스피커가 이에 해당한다. 그리고 자동차 내에서 음성으로 운전자가 운전 중에도 전화를 걸거나, 음악을 변경하거나, 내비게이션 명령을 내려 운전자 안전 운행을 도와준다.

또한 개인비서인 스마트폰은 시리(Siri), 구글 어시스턴트(Google Assistant), 삼성의 빅스비(Bixby)와 같은 인공지능 비서는 사용자의 음성 명령을 인식하여 일상적인 질문에 답하거나, 일정을 관리하고, 알람을 설정하는 등의 작업을 수행한다.

음성으로 텍스트 변환하여 음성 메모를 텍스트로 변환하거나, 문서 작성을 위한 음성 입력이 가능하다. 그리고 기자, 작가, 학생 등 다양한 사용자에게 시간의 효율적인 도움을 주며, 챗봇과 가상 도우미는 고객 서비스에서 질문에 자동으로 답변하거나, 문제를 해결하여 준다.

시각 장애인이나 이동이 어려운 사람들이 기기를 사용하거나 음성 명령을 통해 웹사이트를 탐색하거나 문서를 읽는 데 도움을 주고 있다. 이 밖에도 음성 인식 기술은 의료 분야에서 환자의 상태 모니터링, 교육 분야에서 언어 학습 도구로서, 게임 및 엔터테인먼트 분야에서 사용자 경험 향상 등 다양한 방

107) Mehrish, A., Majumder, N., Bharadwaj, R., Mihalcea, R., & Poria, S. (2023). A review of deep learning techniques for speech processing. Information Fusion, 101869.

면에서 활용되고 있다.

최근에는 생성형 AI 기술을 활용하여 오픈 AI, 메타, 구글 등이 치열한 경쟁을 하고 있다.

[표 2-3] 음성 인식 기술 활용 국내외 사례¹⁰⁸⁾

서비스	출시일	호출명	주요 기기	주요 특징
빅스비 (삼성전자)	2017.3	“하이 빅스비”	- 갤럭시 스마트 폰, 테블릿, 갤럭시 워치 - 스마트 TV, 냉장고	- 빅데이터, 딥러닝 기반 - 스마트 기기, 가전제품 탑재 및 타사제품 개방 계획 - 음성, 카메라 촬영 기능
Q보이스 (LG전자)	2012.7	“하이엘지”	스마트폰, 스마트TV	- 스마트 기기 탑재 - 타 음성 AI 플랫폼 상호보완
누구 (SK텔레콤)	2016.9	“아리아”	스마트 스피커, 셋톱박스	- 멜론, 플로 연동 음악 감상 - 오픈 플랫폼 ‘누구 디벨로퍼스’ 공개, 개방형 생태계 - 음성AI 셋톱박스 탑재
기가지니 (KT)	2017.1	“기가지니”	스마트 스피커, 셋톱박스	- 음성명령 채널이동, 프로그램 검색 등 제어 기능 ‘기가지니 인사이드’ 공개 개방형 생태계 조성
클로바 (네이버)	2017.5	“헤이 클로바”	스마트 스피커	- 인터넷 서비스와 연동 검색, 외국어 번역 - 바이브,지니뮤직, 벅스 등 복수 음악 스트리밍 서비스
카카오i (카카오)	2017.7	“헤이 카카오”	- 스마트 스피커 - 카 인포테인먼트먼트 시스템	- 다음검색,,카톡 메시지 전송 연계 기능, 카카오 네비적용 - 멜론과 연계한 음악 재생
아마존 알렉사	2014.11	“알렉사”	- 스마트 스피커 - 카 인포테인먼트 - 스마트 TV	- 아마존 에코 탑재 - 서드파티 기업의 서비스
구글 어시스턴스	2016.5	“헤이구글” “오케이 구글”	- 안드로이드폰, 테블릿 - 스마트 TV, 스피커 - 카 인포테인먼트	- 구글 어시스턴트 AI 비서 - 알렉사 스킴과 유사 - 방대한 사용자 데이터 확보로 응답 성능 우수
애플 시리	2011.10	“헤이시리” “시리야”	- 아이폰, 아이패드, 맥 PC,애플워치 - 스마트 스피커 홈팟	- 최초 음성 AI 비서 플랫폼 - 자사기기 폐쇄형 탑재
MS 코타나	2015.12	“헤이 코타나”	- 윈도 PC - 안드로이드 기기 (앱)	- 음성 AI 비서 서비스 - 안드로이드 OS 기기 탑재 불가며, 앱설치 방식
바이두 두어	2017.11	“샤오두,샤오두”	- 스마트 스피커 - 서드파티 스마트 스피커 - 카 인포테인먼트	- AI 기반 음성인식과 검색 두어OS 개방형 플랫폼 97% 이상 음성인식 성능 2018년 8월 기준 OS 설치

108) 전창익,이효정,김기범. (2020). 음성 AI시장의 동향과 비즈니스 기회. 삼정KPMG 경제연구원 ISSUE MONITOR, 제126호.

대규모 언어모델(LLM)을 이용한 주요 서비스는 [표 2-4]와 같다.¹⁰⁹⁾

[표 2-4] 대규모 언어모델(LLM) 주요 서비스

구 분	오픈 AI	메타	구글
초거대 언어 모델(LLM)	GPT 시리즈 (2022년 12월 공개) 최근 출시GPT-4는 이미지 인식 가능	라마(LLaMA) 시리즈 (2023년 2월 공개) 최근 오픈소스 라마3 공개	제미니(람다(LaMDA)) (2023년 12월 공개) 멀티모달 AI로 텍스트, 이미지, 음성, 영상을 상호 작용
파라미터 수	100조개 예상(GPT4.0)	4000억개 학습중	1350억개
주요서비스	대화형 AI 챗봇 '챗 GPT' (친화적 상호작용)	대화형 AI 챗봇 '메타 AI'(텍스트 입력으로 작동하는 AI버서)	대화형 AI 챗봇 '바드'(구글 엔진의 정보를 기반 최신 정보 제공)
기타	마이크로소프트 연합전선 파이2 출시(파라미터 7억개), 경량언어모델(sLLM)의 멀티모달 AI로 진화중	바이두 어니 3.0 타이탄 파라미터 2600억개 대화형 생성 AI에 적용 예정	네이버 하이퍼클로바 파라미터 2040억개 한국판 챗 GPT '서치GPT' 출시예정

이처럼 음성 인식 기술은 초거대 생성 AI 기술과 융합한 인간의 명령을 이해하고 지능적이며 사회적인 상호 작용, 인간 증강 기능에서 상호작용을 자연스럽게 교류하는 서비스 로봇의 등장이 가속화되고 있다.

음성 인식 기능이 탑재한 가정용 로봇, 의료 분야에서 환자 모니터링, 기본 의료 지원, 환자 돌봄 서비스가 가능하다. 그리고 호텔, 공항, 쇼핑몰 등에서 고객을 안내하는 서비스를 제공하고, 교육 분야에서도 학습자와 상호작용을 하여 학습자의 흥미와 참여를 증가시킬 것이다.

카트너에서 2030년까지 음성 인식 기술 스마트 로봇과 상호 교류를 인간과 매일 80%에 도달할 것이라고 본다. 음성 인식 기술은 놀라운 속도로 다양한 분야 산업 및 생활 분야에서 혁신을 가져오며, 인식률, 정확도, 실시간 처리, 시맨틱 기반 인식 및 개인화 서비스를¹¹⁰⁾ 통해 상호작용 기술이 확장되어 생활의 편리 및 풍요로움을 만들어 줄 것으로 기대된다.

109) <https://haninpost.com/archives/89763>

110) <https://ongaemi01.com/entry/AI-음성인식-기술의-발전과-적용분야#recentComments>

제 3 절 멀티모달을 활용한 상호작용 연구

1) 멀티모달 시스템 개요

멀티모달은 모달이란 ‘모달리티’에 해당하며 어떤 형태로 나타나는 현상이나 그것을 받아들이는 방식을 뜻하며 ‘멀티모달 인터페이스(Multimodal Interface)’를 줄인 말로 사람과 기계 간의 통신을 위해 음성, 키보드, 펜 등을 이용해 정보를 주고받는 것을 말한다.¹¹¹⁾¹¹²⁾¹¹³⁾ 모달리티는 [표 2-5]와 같이 개별적인 개체에 따라 출력 모달리티, 입력 모달리티로 구분하며, 사람과 기계를 비교하여 각각 모달리티 개념을 이해할 수 있다.¹¹⁴⁾

[표2-5] 사람과 기계간 입출력 모달리티

구 분	오픈 AI	메타	구글
초거대 언어 모델(LLM)	GPT 시리즈 (2022년 12월 공개) 최근 출시GPT-4는 이미지 인식 가능	라마(LLaMA) 시리즈 (2023년 2월 공개) 최근 오픈소스 라마3 공개	제미나이(람다(LaMDA) (2023년 12월 공개) 멀티모달 AI로 텍스트, 이미지, 음성, 영상을 상호 작용
파라미터 수	100조개 예상(GPT4.0)	4000억개 학습중	1350억개
주요서비스	대화형 AI 챗봇 ‘챗 GPT’ (친화적 상호작용)	대화형 AI 챗봇 ‘메타 AI’(텍스트 입력으로 작동하는 AI비서)	대화형 AI 챗봇 ‘바드’(구글 엔진의 기반 최신 정보 제공)
기타	마이크로소프트 연합전선 파이2 출시(파라미터 27억개), sLLM 멀티모달 AI로 진화	바이두 어니 3.0 타이탄 파라미터 2600억개 대화형 생성 AI 적용	네이버, 하이퍼클로바 파라미터 2040억개 한국판 챗 GPT ‘서치GPT’ 출시예정

111) <https://inlpcenter.org/nlp-submodalities-courtney/>

112) 조운상, 이지운, 이영재, 강현규, 김성범. (2019). Multimodal Deep Learning for Explainable Product Design Data. 『대한산업공학회 추계학술대회 논문집』.

113) <https://disquiet.io/@dmchoi/makerlog/ai에서-멀티모달-이-무엇이고-왜-중요할까>

114) Gi-Hyeong, H. (2006). XML 기반 멀티모달 상호작용 표준 언어. Korea Multimedia Society, 10(1), 71-82.

멀티모달 상호작용은 1개 이상 모달리티를 사용하는 것으로 모달리티 결합을 통한 시너지 효과가 있고, 각 모달리티 별 장점을 극대화하고, 사용자 선택의 자율성과 커뮤니티의 자연스러운 인터페이스를 가능하게 한다.¹¹⁵⁾ 예를 들어 컴퓨터상에 마우스로 특정 아이콘을 선택하고 키보드로 텍스트를 입력하는 일반적인 방식에서 손동작 인식으로 특정 아이콘을 지정하고 음성으로 텍스트를 입력하면 더욱 편리하게 컴퓨터를 다룰 수 있다. 멀티모달의 상호작용은 [표 2-6]과 같이 인터페이스 종류에 따라 기본과 확장으로 구분하여 다양한 종류 및 특징으로 나눌 수 있다.

[표 2-6] 멀티모달 인터페이스 종류

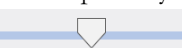
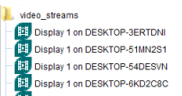
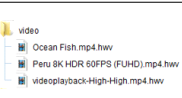
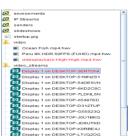
구분	종 류	특 징	기 술
기본	음성 인터페이스	음성 입력→기계 입력 기계 출력→음성 출력	음성인식, 음성합성
	키보드 및 마우스	키보드 및 마우스 인식	키보드 마우스 인식
	잉크 인터페이스	펜 글씨 및 그림 표시 기능	펜 필체 인식
확장	NUI (Natural User Interface)	제어장치 없이 감각, 행동인식 능력 직접 교감	디지털 기기 제어
	TUI (Touch User Interface)	실물 이용 디지털 정보와 상호 구현	체감형 인터페이스
	웨어러블 햅틱	착용감, 항시성, 안정성, 사회성 반영 착용형 컴퓨팅	HMD, 포스 터치, 모션 인식
	HRI (Human Robot Interface)	로봇과 인간간 소통	얼굴인식, 제스처인식, 음성인식
	감성 인식	사용자 시각, 청각, 촉각, 자극 모바일 전자기기와 교감	패더스, 컴저트 붓
	음성 인식	마이크, 전화이용 음성신호를 텍스트로 변환 인식	음성인식, 보이스 인식 기술
	제스처 인식	카메라 및 키넥트, 레이저, 라이더로 손동작이나 사람의 신체 인식	제스처 인식 기술

115) Benoit, C., Martin, J. C., Pelachaud, C., Schomaker, L., & Suhm, B. (2000).

Audio-visual and multimodal speech systems. Handbook of Standards and Resources for Spoken Language Systems-Supplement, 500, 1-95.

본 연구에서는 손동작과 음성 인식으로 비디오월 제어시스템을 어느 정도 제어할 수 있는지 집중적으로 분석하였고, 손동작으로 제어가 어려운 부분을 음성으로 대체가 가능 정도를 살펴보았다. H사의 비디오월 제어시스템을 기준으로 주요 기능별로 구분하여 손동작과 음성 인식으로 제어 가능한 기능과 기술적으로 적용 여부를 나눠 [표 2-7]과 같이 표시하여 비교하였다.

[표 2-7] 손동작 및 음성인식 적용 가능성 비교

제어 UI 컴포넌트	제어 아이콘	기 능	가능한 인식
슬라이드 컨트롤 (Slade Controls)	Object Zoom 	• 선택된 콘텐츠 확대 축소 배율 조정	• 기존 마우스 방식 제어
	Transparency 	• 콘텐츠 투명도 제어	
	Rotation 	• 콘텐츠의 회전 조절	
드롭다운 (Dropdown)	video_streams 	• 동영상 저장 폴더	• 손동작 인식, 음성 인식, 마우스 인식
	video 	• CCTV 저장 폴더	• 손동작 인식, 음성 인식, 마우스 인식
스크롤바 (Scrollbar)		• 위/아래 메뉴 표시	• 손동작 인식, 음성 인식, 마우스 인식

[표 2-7]과 같이 제어 UI 컴포넌트 별로 슬라이드 컨트롤, 드롭다운, 스크롤 바로 구분하여 제어 아이콘의 메뉴를 손동작 및 음성 인식 제어가 가능한 지 비교하여 설계에 반영하여 구현 및 검증을 통해 효용성을 연구하였다.

슬라이드 컨트롤은 슬라이드 바의 좌우로 움직여 콘텐츠를 확대하고 축소하는 기능, 콘텐츠의 투명도, 콘텐츠 회전 조절 등은 마우스 좌표 인식 구현이 어려움이 있어 본 연구에서는 제외하였다.

콘텐츠 확대 및 축소, 콘텐츠의 표출 등을 가장 많이 사용하는 기능으로 디스플레이의 표출화면에서 손동작과 음성 인식으로 제어되도록 기능을 추가하

였다. 비디오월 제어 기능 전체를 손동작 및 음성 인식으로 제어하는 것은 현재 기술적인 한계로 일부 유용한 기능을 적용하여 멀티모달의 기능의 편리성과 유용성을 검증하는 데 중점을 두고자 한다.

2) 멀티모달 시스템의 활용 사례

멀티 모달 시스템은 다양한 분야와 산업에서 적용되며, 스마트홈 및 IoT 기기 제어에서 음성, 제스처, 터치 등에서 다양한 입력 모달리티를 활용하여 스마트홈 및 IoT 기기를 편리하게 제어하는 것은 물론 차량 내에 다양한 센서 데이터(영상, 음성, 운전자 제스처)를 종합적으로 분석하여 운전자의 상황을 인지하고 대응하며, 환자의 다양한 생체 정보(영상, 음성, 생체신호 등)를 분석과 환자의 자연스러운 의사소통을 위한 음성, 제스처 등을 다양한 입력방식을 통해 질병 진단 및 치료를 계획한다.¹¹⁶⁾ 그리고 교육 및 훈련 분야에서도 학습자의 반응, 감정, 집중도 등을 다양한 센서 데이터를 통해 개인화된 학습 경험을 높이고 있다.¹¹⁷⁾



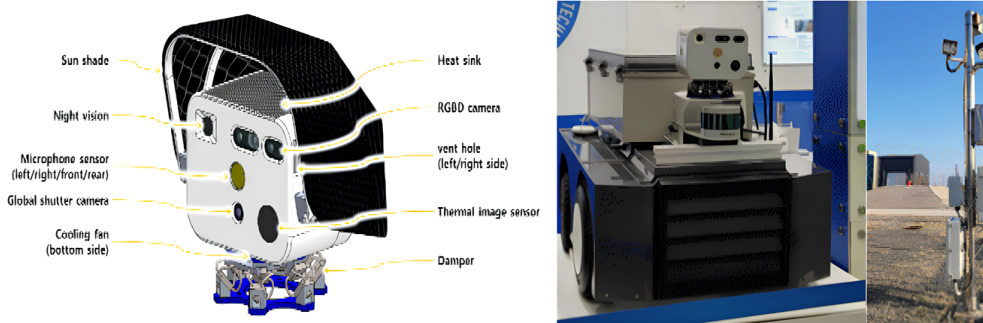
[그림 2-8] 멀티모달 기반 스마트홈 AI 에이전트

[그림 2-8]과 같이 LG 전자에서 출시한 ‘LG클로이 서비스 봇’으로 자율주행은 기본이고 음성을 인식하고 응답하거나 챗봇 서비스를 제공하고 사용자 감정까지 파악하는 기능으로 발전하고 있다.¹¹⁸⁾

116) <https://www.apple-economy.com/news/articleView.html?idxno=70290>

117) <https://www.ewestoday.co.kr/news/articleView.html?idxno=2047564>

여러 대의 센서와 카메라를 부착하여 건물 밖의 시설에 대해 감시하고 순찰하는 로봇도 등장하여 장애물이나 좁은 공간 및 통로를 구분하여 순찰하고, 방문자의 얼굴을 인식하여 침입자를 회사관계자에게 통보를 해주는 멀티모달 센서 기반 경비 로봇이 서비스되고 있다.¹¹⁹⁾



[그림 2-9] 멀티 모달 센서 경비 로봇

현재 사용자를 중심으로 다양한 멀티모달 방법들이 연구되고 있고, 제스처 인식 기술을 이용한 음성 인식 기술 등 다양한 기술과 융합하여 더 편리하고 효율적인 인터페이스를 제공한다.¹²⁰⁾

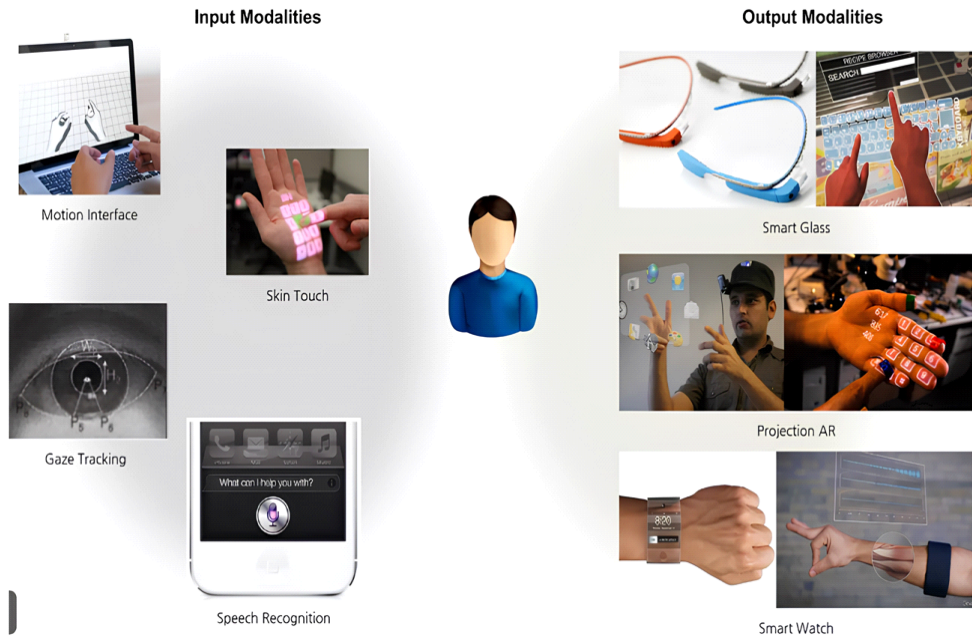
[그림 2-10]과 같이 입력 모달리티에서 손동작, 아이 트래킹, 스킨 터치, 음성인식으로 구성되어 있고, 출력 모달리티는 스마트 글라스, 프로젝션 AR, 스마트 와치 등으로 구현되고 있다. 추가로 휴먼케어 로봇, 협동 로봇과 상호 작용에 대한 부분도 출력 모달리티에 해당된다.

향후 NUI(Natural User Interface) 융합 플랫폼의 연구가 진행되어 인간이 인식하고 있는 모든 감각 기관과 컴퓨터 간의 상호작용이 가능한 시스템으로 발전할 것으로 전망된다. 멀티모달의 영상, 음성, 텍스트, 위치 정보 등의 다양한 데이터를 융합하여 상황을 종합적으로 분석하여 자동으로 의사결정을 할 수 있도록 관리자에게 전달될 것이며, 관리자는 사용자 친화적인 손동작과 음성 인식을 통해 다양한 정보를 쉽고 빠르게, 편리하게 모니터링하여, 더 나은

118) <https://ddaily.co.kr/m/page/view/2023122915351101186>

119) <https://contents.premium.naver.com/edumeta/edu/contents/230718024326513mx>

120) <https://www.slideshare.net/JonghoonSeo/20-24091955>



[그림 2-10] 멀티 모달 간의 융합 사례

환경에서 안정성과 사용자 경험을 높일 수 있을 것이다.

제 3 장 멀티모달 기반 비디오월 제어시스템

상호 작용 설계

제 1 절 멀티모달 기반 비디오월 제어시스템의 상호작용 정의

1) 비디오월의 일반적인 정의

비디오월이란 여러 대의 컴퓨터 모니터, 비디오 프로젝터 또는 TV를 나란히 놓거나 서로 쌓아 하나의 대형 화면을 형성하는 특수한 다중 모니터 설정이라 정의하며, 일반적인 디스플레이 기술에는 LCD 패널, 다이렉트 뷰 LED 어레이, 혼합 프로젝션 스크린, 레이저 인광체 디스플레이 및 후면 프로젝션 큐브가 포함된다.¹²¹⁾

비디오월은 일반적으로 LCD 패널, 리어 프로젝션 큐브, 다이렉트 LED 타일을 이용한 방법으로 사용되고, 최근에는 LED 타일을 사용한 해상도를 높이면서 대형 외장 전광판, 실내용 사이니지의 형태로 사용되거나 버추얼 프로젝션이나 통합 관제센터에 활발하게 사용이 되고 있다.¹²²⁾

그리고 비디오월은 하나의 논리적인 화면을 형성하기 위해 함께 고정된 둘 이상의 비디오 화면으로 구성된 대형 디스플레이며, LCD, LED 컴퓨터 화면, 기존 텔레비전 세트 및 비디오 프로젝터를 사용하여 형성할 수 있다.¹²³⁾

또한 공공기관의 스마트 통합관제센터, 방송국 XR 스튜디오, 철도 스마트 통합센터, 발전소의 상황을 모니터링하는 상황실, 스마트 제조공장의 제조과정을 모니터링하는 관제센터, 물류 센터의 물류 흐름을 파악하는 관제센터, 실외 옥외 광고 전광판 등에 다양한 용도로 사용하고 있는 공간의 가치를 높이는 시스템으로 자리 잡고 있다.

121) <https://academic-accelerator.com/encyclopedia/kr/video-wall>

122) 박강우, 하동환. (2022). LCD 모니터를 활용한 버추얼 스튜디오용 다중 모니터 구현 방법에 관한 연구. 『한국과학예술융합학회』, 40(5), 119-128, 10.17548/ksaf.2022.12.30.119

123) <https://ko.isemc.com/news-isemc/what-is-video-wall>

2) 비디오월 제어시스템 정의

비디오월 제어시스템은 비디오월(출력)과 연결된 모든 입력 소스를 관리할 수 있는 장치로 다양한 정보와 데이터를 시각화하기 위한 하드웨어와 소프트웨어로 구성되어 비디오월 시스템의 디스플레이 및 작동을 관리하고 제어하는 시스템이라 정의한다.¹²⁴⁾ 비디오월 제어는 시대별로 1세대 Matrix Switcher 방식, 2세대 서버형 하드웨어 기반 비디오월 제어 방식, 3세대 소프트웨어 기반 AII IP 비디오월 제어 방식으로 나뉘 발전해 왔다.¹²⁵⁾

비디오월 제어시스템은 영상 신호를 확장하거나 분리, 조합하는 하드웨어를 이용하는 비디오월에서 영상 압축 기술과 IP 네트워크를 이용하여 여러 영상을 동시에 재생하거나 표시하는 시스템이다.¹²⁶⁾ 본 연구는 소프트웨어 기반 비디오월 제어시스템을 중심으로 비디오월의 상호작용 연구를 수행하였다.

3) 멀티모달 인터페이스의 기능적 정의

제1장 3절에 언급한 바와 같이 멀티모달 인터페이스에서 모달이라는 개념은 모달리티(Modality)를 뜻하며¹²⁷⁾, ‘모달리티란 인터랙 커뮤니케이션 채널이다’ 라도 정의한다.¹²⁸⁾ 멀티모달 인터페이스는 인간의 제스처, 시선, 손의 움직임, 행동의 패턴, 음성, 물리적인 위치 등 인간의 자연스러운 행동들에 대한 정보를 해석하고 부호화하는 인지 기반 기술이다.

인간과 컴퓨터 인터랙션 과정에서 동시 여러 모달리티의 입출력을 허용하며 다수의 모달리티의 조합과 입력 신호 통합해석 등을 통해서 상호 의사를

124) <https://ko.isemc.com/news-isemc/what-is-a-video-wall-controller>

125) 조영욱, 노광현. (2023). IP 비디오 Wall 기반 실시간 양방향 모바일 방송시스템에 관한 연구. 『아시아태평양융합연구교류논문지』, 9(2), 1-12.

126) 양선진, 박재표, & 양승민. (2019). 스마트 시티 환경에서 대규모 영상 모니터링을 위한 IP 비디오 월 시스템의 설계 및 구현. 『한국산학기술학회 논문지』, 20(9), 7-13.

127) <https://m.blog.naver.com/vinylx/220367088474>

128) <http://www.w3.org/TR/2003/NOTE-mmi-reqs-20030108/>

교환하는 것이다.¹²⁹⁾

[표 3-1]과 같이 주요 인터페이스별 기능 설명을 통해 다양한 인터페이스가 기능 활용도에 따라 서비스되고 있다.

[표3-1] 주요 인터페이스 기능

No	주요 인터페이스	기능 설명	예시
1	음성 인식	음성 인식 명령 수행, 질문 응답	아마존 알렉사, 구글 어시스턴트
2	제스처 인식	손동작 및 신체 동작 인식 명령 수행	닌텐도 Wii, 마이크로소프트 Kinect
3	터치 인식	터치스크린 화면 명령 수행	스마트폰, 태블릿, 터치스크린, 키오스크
4	시각 인식	이미지 인식 및 분석	얼굴 인식 시스템, QR 코드 스캐너, 자율 주행 자동차
5	텍스트 입력	키보드, 터치스크린 입력	컴퓨터 키보드, 스마트폰 텍스트 입력
6	멀티모달 인식	여러 입력방식 결합 명령 수행	음성과 제스처로 컴퓨터 제어
7	햅틱 인식	촉각 피드백 기능 제공	스마트폰 진동 알림, VR 장비의 촉각 피드백
8	컨텍스트 인식	사용자의 상황, 환경 인식 적절한 반응 제공	위치 기반 서비스, 상황맞춘 개인 정보 제공

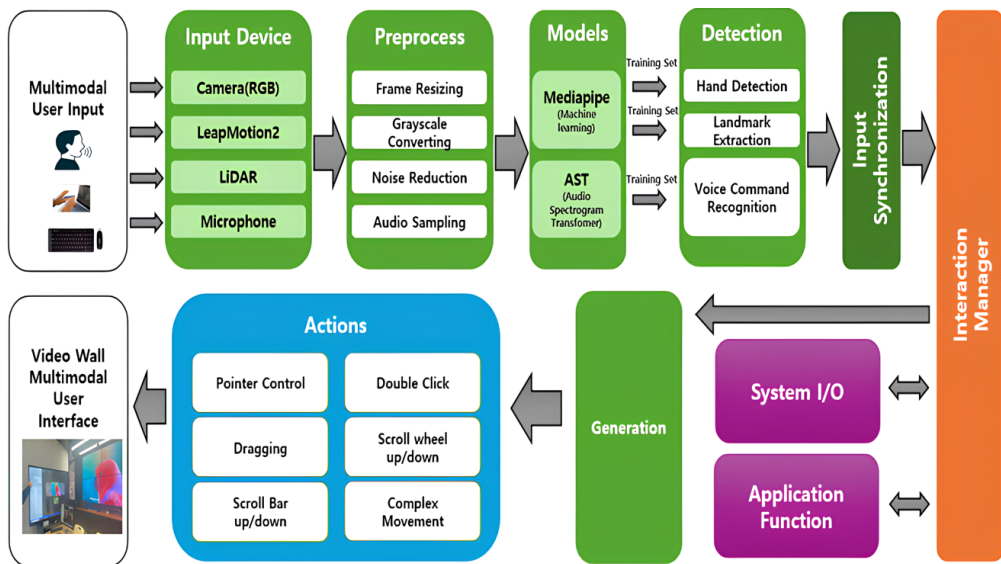
본 연구에서는 손동작과 음성 인식을 기반한 비디오월 제어시스템을 멀티모달 인식으로 제어하여 기존 키보드와 마우스 운용의 단점을 보완하고, 사용자의 유용성을 높이는 연구이다. 스마트 시티의 수 많은 CCTV 영상을 제어하는 것은 단순한 일이 아니며 통합센터 운영하는 관제사의 업무 한계를 넘어선 지 오래되었다. 멀티모달 인터페이스의 기능을 적용한 통합관제센터는 관제사의 업무를 보조하여 업무의 효율을 높이는 것은 물론 중요한 기능으로 역할을 할 것이다.

129) 임미정, & 박범. (2006). 멀티모달 인터랙션을 위한 사용자 병렬 모달리티 입력방식 및 입력 동기화 방법 설계. 『대한인간공학회지』, 25(2), 135-146.

제 2 절 멀티모달 기반의 비디오월 제어시스템 아키텍처

1) 비디오월 제어시스템의 상호작용 체계

본 연구에서는 차세대 비디오월 제어시스템 상호작용을 위한 아키텍처를 [그림 3-1]과 같이 제안한다. 비디오월 제어시스템의 기존 상호작용 인터페이스 방식인 키보드&마우스에 손동작과 음성 인식을 추가하는 운영할 수 있는 구조이다. 손동작 인식하기 위해 웹 카메라(RGB), 3D 카메라, Leap Motion, LiDAR 등의 인터페이스 장비를 사용할 수 있는데 본 연구에서는 손동작 인식을 위한 인터페이스 장치는 가격이 저렴하고 널리 보급된 웹 카메라(RGB)를 활용하고, 음성 인식을 위한 인터페이스 장치는 마이크를 활용하였다.



[그림 3-1] 멀티모달 기반 비디오월 제어 아키텍처

입력 장치에서 손동작 및 음성은 전처리 과정을 거쳐 정의된 동작 라이브러리를 활용하여 모델링을 거쳐 실제 Detection 검증 후 입력 명령을 동기화하여 비디오월 제어시스템에 상호작용이 가능하도록 설계되었다. 비디오월 제

어시시스템에서 사용 빈도가 높은 제어 기능을 정의하여 손동작 및 음성 인식이 적용될 경우 동작의 우선순위를 적용하여 제어하였다.

본 연구의 동기화는 순차적으로 손동작을 인식한 후 음성으로 제어가 되며, 먼저 손동작으로 마우스를 이동하여 해당 아이콘을 지정하고 음성 명령을 내려 제어하도록 설계하였고, 두 동작이 동시에 작동할 때에 콘텐츠 명령의 혼선을 최소화하도록 순차적인 동기화 방식으로 구현하였다.

2) 입력 모달리티 전처리 및 모델링

손동작 인식 전처리에서 웹 카메라를 통해 사용자의 손 움직임을 실시간으로 캡처하여 손영역을 분리하여 손영역의 관절 위치, 손가락 각도, 손 모양 등을 추출하여 손동작을 인식하였다. 이와 같은 동작의 특징을 바탕으로 사전에 정의된 라이브러리 모델인 MediaPipe를 통해 손동작 인식과 비교하여 마우스의 기능인 마우스 포인트 이동, 더블 클릭, 좌클릭, 우클릭, 휠 업/다운(확대/축소)을 손동작을 검증하였다.¹³⁰⁾

음성인식은 명령어에 대한 전처리 과정에서 노이즈 리덕션(Noise Reduction)과 오디오 샘플(Audio Sampling)을 통해 정의된 모델링인 AST(Audio Spectrogram Transformer)를 활용하여 알고리즘을 구성하였다.¹³¹⁾ AST(Audio Spectrogram Transformer)는 ViT(Vision Transformer)와 동일하지만 오디오에 적용하는 모델로 오디오는 이미지(스펙트로그램)로 변환된 후 Vision Transformer가 적용되었다.

노이즈 리덕션은 오디오나 비디오 등 다양한 미디어 데이터에서 불필요한 잡음을 제거하여 품질을 개선하는 것이다. 오디오 노이즈 리덕션(Audio Noise Reduction)의 방법은 스펙트럼 감산법(Spectral Subtraction), 스펙트럴 게이팅(Spectral Gating), 노이즈 프로파일링(Noise Profiling), 위너 필터(Wiener Filter), 노이즈 게이트(Noise Gate) 등이 방법이 있다.¹³²⁾

130) <https://giveme-happyending.tistory.com/202>.

131) <https://velog.io/@fbdp1202/Sound-Classification-정리.-AST-Audio-Spectrogram-Transformer>

132) https://en.wikipedia.org/wiki/Noise_reduction

그리고 이미지 및 비디오 노이즈 리덕션(Image and Video Noise Reduction)의 방법은 평균 필터(Averaging Filter), 미디언 필터(Median Filter), 가우시안 블러(Gaussian Blur), 비지역 평균(Non-Local Means), 웨이블릿 변환(Wavelet Transform), 주파수 도메인 필터링(Frequency Domain Filtering) 등의 방식을 사용한다.¹³³⁾

최근 많이 사용되고 있는 스펙트럼 감산법(Spectral Subtraction)과 Deep Learning 기반의 방법을 도입하였다.

3) 입력 동기화와 인터랙션 매니저

멀티모달 입력 동기화 형식은 시간 기반 동기화(Temporal Synchronization), 공간 기반 동기화(Spatial Synchronization), 의미 기반 동기화(Semantic Synchronization), 하이브리드 동기화(Hybrid Synchronization)로 구분한다.¹³⁴⁾

본 아키텍처의 입력 동기화 방법은 하이브리드 동기화 (Hybrid Synchronization)의 시간과 공간, 의미를 모두 고려한 다중 기준 동기화를 채택하여 사용자가 손동작을 특정 위치에서 관련 음성명령을 내리면 명령 의미를 파악하여 최종 동작을 결정하는 방식이다.

다음으로 인터랙션 매니저는 다양한 입력 모달리티를 관리하여 사용자와 제어시스템 간의 상호작용을 원활하게 하는 역할을 한다. 여러 입력 모달리티가 충돌하거나 모순되는 경우, 우선순위를 정하여 명령을 수행하도록 하였다. 이와 같이 순차적으로 정의된 입력 모달리티에서 최종 결정된 명령은 일반 콘텐츠 명령 수행으로 전달되어 제어화면의 콘텐츠가 디스플레이 표출이 되는 구조로 설계하였다.

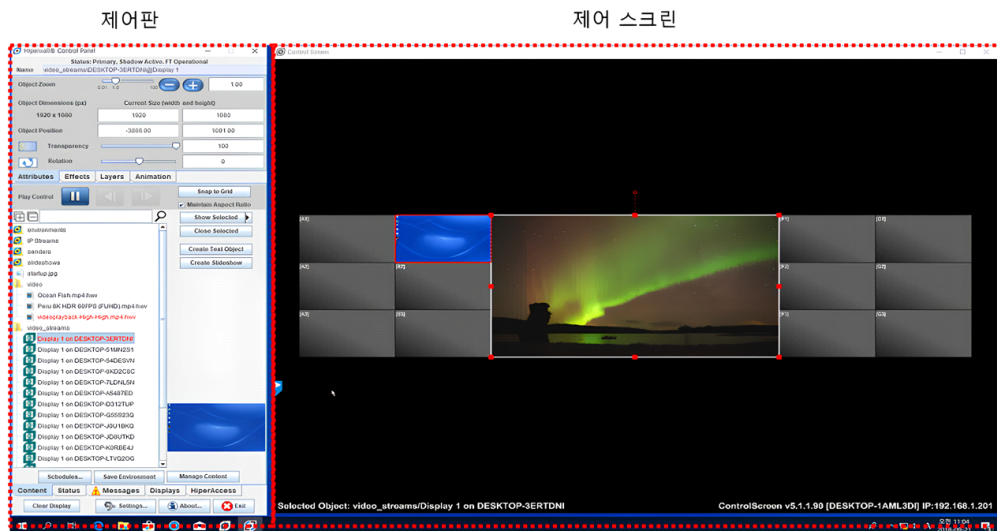
133) Kim, Y.-H. and Nam, J.-H. (2011) "Estimation of the Noise Variance in Image and Noise Reduction," Korean Journal of Applied Statistics. The Korean Statistical Society. doi: 10.5351/kjas.2011.24.5.905.

134) Guo, W., Wang, J., & Wang, S. (2019). Deep multimodal representation learning: A survey. Ieee Access, 7, 63373-63394.

제 3 절 멀티모달 기반의 비디오월 제어시스템 설계

1) 손동작 및 음성 인식으로 비디오월 제어시스템의 기능 정의

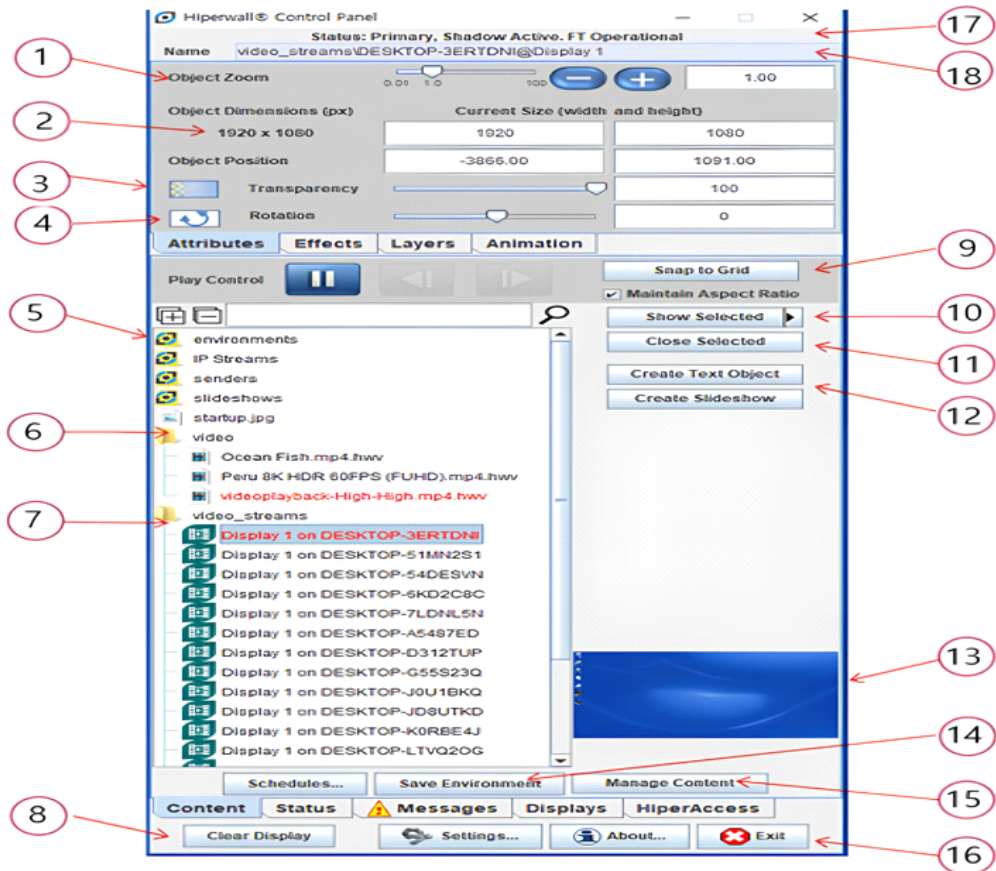
본 연구에서 H사의 비디오월 제어시스템을 중심으로 기능을 분류하고, 그 기능에 대해 손동작 인식으로 비디오월 제어시스템의 상호작용을 설계하였다. 먼저 H사의 비디오월 기능은 아래 표와 같이 가장 많이 사용하는 기능을 세분화하고 그 기능을 손동작 및 음성 인식으로 제어가 가능한 기능을 구분하였다.



[그림 3-2] H사의 비디오월 제어시스템 UI화면

[그림 3-2]과 같이 제어판 부분은 비디오월에 표출되는 모든 화면을 제어하고 조정하는 UI(User Interface)화면이고, 제어 스크린 부분은 비디오월과 동일한 모양으로 음영 이미지를 나타내며 제어판에서 컨트롤한 화면이 비디오월과 동일하게 표출이 되는 UI 화면이다. [그림 3-2]는 [그림3-1]의 제어판을 확대하여 설명한 그림으로 자주 사용하는 기능을 나눠 표시하였다. 제어판 기능은 비디오월 제어하는데 핵심 기능들로 구성되어 있고 손동작과 음성 인식하여 제어판의 기능을 제어함으로써 손동작 및 음성으로 제어 스크린을 통해 비디오월에 상호작용이 되도록 구성되었다.

[그림3-3]는 자주 사용하는 기능을 표시한 것으로 1번부터 18번으로 기능을 마우스 기능인 콘텐츠 이동, 스크롤 기능, 콘텐츠 축소/확대 기능을 참조하여 비디오 월 제어판에서 기능이 손동작 및 음성 인식으로 제어할 각 기능별로 적용 여부 및 기존 마우스로 제어가 더 쉬운 부분을 구분하여 설계하였다. [표3-1]은 [그림 3-3]의 기능을 설명하고 손동작 및 음성 인식으로 제어가 예상되는 부분을 나타내고 있다.



[그림 3-3] H사의 비디오월 ‘A’ 제어판 UI화면

[표 3-2]은 H사의 비디오월 제어판 UI 화면의 기능 설명과 이 기능에 대한 손동작 활용 가능 여부를 설명한 것으로 손동작으로 마우스 이동, 클릭, 화면 확대, 축소 등을 가능한 부분을 고려하여 설계를 진행하였다.

[표3-2] H사의 비디오월 제어판 UI 화면 기능 및 동작

No	기 능	설 명	동작 적용 여부
1	Object Zoom	선택한 콘텐츠의 확대 축소 배율 표시	기존 마우스 적용
2	Object Dimentions	콘텐츠 해상도 표출	적용과 관계 없음
3	Transparency	콘텐츠 투명도 조절	마우스 적용
4	Rotation	콘텐츠 회전 및 실시간 회전각 표출	마우스 적용
5	Environment	기존 환경 저장 및 불러오기	손동작, 음성인식
6	Video	동영상 저장 폴더	손동작, 음성인식
7	Video Streams	cctv 화면 저장 폴더	손동작, 음성인식
8	Clear Display	모든 화면 일시에 삭제	손동작, 음성인식
9	Snap to Grid	비디오월에 표출되는 콘텐츠 해상도 모드 - 체크시: 원본 콘텐츠 유지 표출 - 해제시: 원본해상도 관계없이 표출	손동작 인식
10	Show Selected	실시간 선택된 콘텐츠만 복사	손동작 인식
11	Closed Selected	실시간 선택된 콘텐츠만 삭제	손동작 인식
12	CreateText Object	텍스트 제작할 때 사용 아이콘	손동작 인식
13	실시간 표출화면	제어판의 화면 미리보기 - 한번 클릭: 미리보기 - 더블 클릭: 비디오 업로드 (화면 표출)	손동작 인식
14	Save Environment	운영자 환경에 저장	손동작 인식
15	Manage Content	제어판 화면을 생성 및 삭제하는 기능	손동작, 음성인식
16	Exit	시스템 종료	손동작 인식
17	Status	주,예비시스템 동작 상태 표시	기존 마우스 적용
18	Name	제어판에 선택된 콘텐츠 파일명 표시	기존 마우스 적용

[표 3-2]과 같이 손동작으로 제어가 가능한 부분 중에 음성 인식을 적용하여 제어가 가능한 기능을 분석해 보면 [그림 3-3]에서 제어판 기능 UI로 음성으로 마우스를 포인트를 이동하는 부분은 마우스 포인트의 움직임

임의 위치와 궤적 정보를 구현의 한계가 있어 마우스 이동하는 부분은 손 동작으로 구현하고 더블 클릭이나 좌클릭, 우클릭, 스크롤 다운/업, 마우스 휠의 업(확대)/다운(축소)을 음성으로 제어하는 기능을 설계하였다.

[그림 3-3] 제어판 메뉴 UI 컴포넌트 중 Object Zoom, Transparency, Rotation의 슬라이더 컨트롤은 슬라이더의 특정 범위의 값을 선택할 때 자주 사용되는 UI로, 슬라이더를 손가락 터치 또는 마우스로 드래그함으로써 미세하게 값을 조정할 수 있고 콘텐츠의 확대/축소, 투명정도, 회전 등을 조절할 때 사용한다.

[표 3-3] 손동작 및 음성 인식 기반 비디오월 제어판 기능 정의

비디오월 제어판 기능 정의	기존 방식	본 연구 방식	기능 측면
제어판 폴더 제어	마우스 이동 및 더블클릭	폴더 이동 : 손동작 인식 더블클릭 : 음성 인식	유용성, 정확성
제어판 스크롤바 제어	마우스 스크롤	스크롤바 이동: 손동작 인식 스크롤바 제어: 음성 인식	유용성, 정확성
제어스크린 콘텐츠 확대	마우스 휠업	콘텐츠 이동 : 손동작 인식 콘텐츠 확대 : 음성 인식	유용성, 정확성
제어스크린 콘텐츠 축소	마우스 휠다운	콘텐츠 이동 : 손동작 인식 콘텐츠 축소 : 음성 인식	유용성, 정확성
제어스크린 콘텐츠 삭제	마우스 좌/우 클릭	콘텐츠 이동 : 손동작 인식 콘텐츠 삭제 : 음성 인식	유용성, 정확성

[표 2-7]에서 언급한 바와 같이 컴퓨터 화면에서 손가락의 움직임, 손 자세 등 다양한 사용자의 손동작을 실시간으로 정확하게 감지하여 슬라이드 컨트롤 UI를 제어하는 것은 구현하는데 기술적인 난이도 문제로 본 연구에서는 기존 방식인 마우스로 제어하는 방식을 적용하였고, 손동작과 음성 인식으로 비디오월 제어판을 제어하는 기능에 추가하여 운영의 유용성과 정확성을 중심으로 설계에 반영하였다. 시스템 기능 정의는 [표 3-3]과 같다.



[그림 3-4] 제어판과 제어스크린 기능별 표출 콘텐츠

[그림 3-4]은 제어판과 제어 스크린에서 5가지 기능을 정의하여 표출되는 콘텐츠가 마우스로 작동하였을 때와 비교하여 사용자 관점에서 유용성 및 정확성을 측정하였다.

- **기능 정의 1** : 컴퓨터의 웹캠을 통해 손동작을 인식하여 커서를 손동작으로 이동하여 해당 폴더위치에 고정하고 ‘two’라고 음성으로 명령을 스피커를 통해 지시하면 폴더의 콘텐츠 파일 이름들이 나타내게 되고 다시 손동작으로 콘텐츠 파일 이름으로 이동하여 다시 ‘two’라고 하면 콘텐츠가 제어 스크린에 표출되며 디스플레이 화면에도 동시에 나타나는 기능이다. 마우스의 기능에서 보면 포인트가 이동하여 폴더 클릭하여 콘텐츠를 더블클릭하면 콘텐츠가 제어 스크린에 표출되는 기능이다.
- **기능 정의 2** : 제어 스크린에 표출된 이미지나 영상을 마우스로 좌클릭하여 콘텐츠를 지정하고 우클릭하면 폴더가 나타나 폴더를 지

정하면 화면이 제거되는 기능을 손동작으로 콘텐츠를 지정하여 ‘two’라고 음성으로 명령을 내리면 콘텐츠가 지정되고 지정된 콘텐츠에서 음성으로 ‘left’와 ‘right’를 명령하면 해당 폴더가 나타나고 다시 손동작으로 이동하여 clear 폴더에 고정하여 음성으로 ‘two’라고 명령을 내리면 지정 콘텐츠가 제거된다.

- 기능 정의 3 : 제어판의 스크롤 바를 손동작으로 커서를 이동하여 스크롤 바 위치에 고정하여 ‘up’, ‘down’을 음성으로 명령하면 스크롤 바가 위와 아래로 이동하는 기능으로 제어판의 표시된 폴더를 위와 아래로 아이콘을 볼 수 있는 기능이다.
- 기능 정의 4, 5 : 는 제어표시판에 표출된 콘텐츠를 확대 및 축소하는 기능으로 손동작으로 제어표시판의 콘텐츠로 이동하고 지정된 콘텐츠를 손동작 크기를 표시하고 음성으로 ‘up’, ‘down’의 명령을 내리면 손동작의 크기만큼 콘텐츠의 화면을 확대 및 축소하는 기능이다.

비디오월 제어시스템의 제어판과 제어표시판에서 손동작이나 음성 인식으로 제어가 되는 기능 정의를 5가지로 한정한 이유는 마우스 기능보다 사용자의 운영 및 효율적인 측면에서 고려하여 기능을 선정하였다. 제어판 및 제어표시판 전체 기능을 손동작이나 음성으로 제어하는 부분은 고도의 기술적 난이도가 요구되므로 본 연구에서는 일부 기능만 적용하여 사용자의 능력을 고려하고 사용자의 처한 상황과 목적을 파악하여 사용자가 제어판의 기능의 세부적인 기능을 자세하게 알지 못해도 사용자가 쉽게 이용할 수 있도록 유용성과 정확성을 검증하기로 하였다.¹³⁵⁾

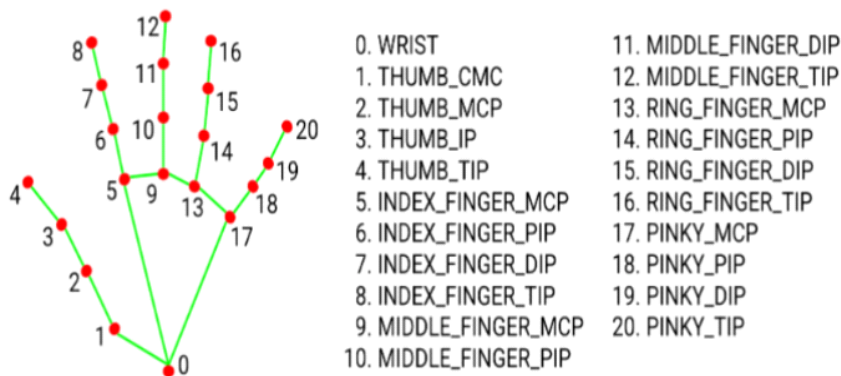
2) 손동작 인식 기반 상호작용 설계

가) 연구내용

본 연구는 비디오월 시스템의 제어 부분을 마우스나 키보드 없이 손동작 인식으로 제어가 되도록 기능을 구분하여, 다양한 기능 중 제한적으로 자주

135) 김고운.(2020). Natural User Interface 표준화 동향 연구. 한국통신학회, 750-752.

사용하는 기능을 기준으로 손동작 인식으로 제어가 되는지를 설계하였다. 손동작 인식은 구글 사의 머신러닝 기반 고성능 손 제스처 인식 솔루션인 MediaPipe Hands를 활용하여 손동작 인식을 구현하였다.¹³⁶⁾ MediaPipe는 오디오 및 비디오 다른 모달리티의 지각 데이터를 처리하기 위한 오픈 소스 크로스 플랫폼 프레임워크이다.¹³⁷⁾ 이는 각 손에 대해 21개의 3D 랜드마크를 프레임마다 추론하여 두 개 이상의 손을 동시에 인식하고 인터페이스가 가능하도록 설계되었다.

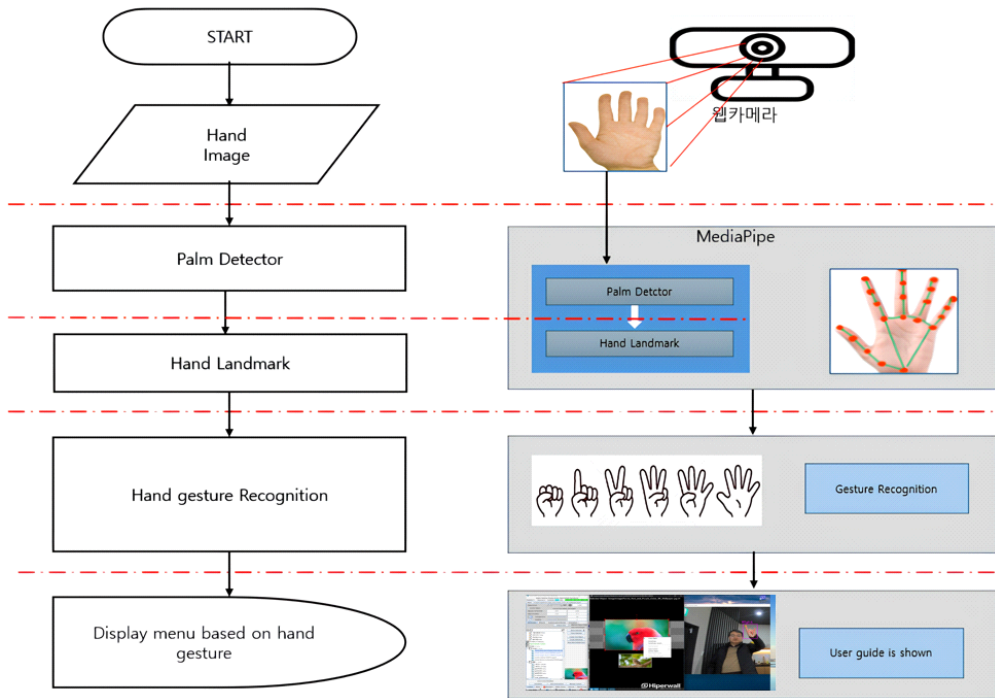


[그림3-5] MediaPipe Hands Landmarks

사용자의 손동작 인식을 기준으로 마우스 이동, 더블클릭, 좌클릭, 우클릭, 스크롤 다운, 스크롤 업, 휠업(확대), 휠다운(축소) 그리고 비디오월 제어시스템에서 제어되는 이미지 및 영상을 생성하거나 삭제하는 기능을 구분하였고, 각각 기능에 따라 손동작 3가지 동작으로 묶어 테스트하여 가장 인식률이 우수하고 신호도가 높은 손동작이 어떤 것인지 추출하였다.

¹³⁶⁾ <https://developers.google.com/mediapipe/solutions>

¹³⁷⁾ Kavana, K. M., & Suma, N. R. (2022). Recognition of Hand Gestures Using Mediapipe Hands. International Research Journal of Modernization in Engineering Technology and Science, 4(06).



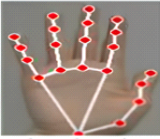
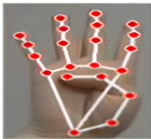
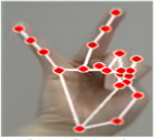
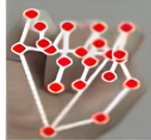
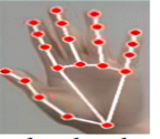
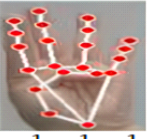
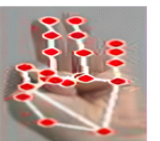
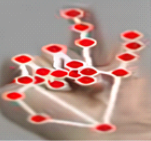
[그림3-6] MediaPipe 활용 손동작 인식 Workflow

[그림 3-6]와 같이 워크플로우의 기반하여 손동작 인식하기위해 필요한 라이브러리들인 OpenCV, cvzone의 HandDetector, mouse, numpy 등이 import되었고, cvzone의 HandDetector 객체는 웹캠 이미지에서 손을 찾는데 사용되었다. 손이 감지되면, 해당 손의 동작에 맞는 기능을 실행하도록 하였다. 마우스와 키보드 등의 사용자 입력을 제어하기 위해 mouse와 pyautogui 라이브러리를 사용하였다. 이들 라이브러리는 마우스와 키보드의 동작을 프로그래밍적으로 제어하도록 해주며, pyautogui는 더불어 화면 캡처, 이미지 인식 등의 기능도 제공한다.

이를 통해 사용자의 입력 없이도 마우스나 키보드를 제어하는 프로그램을 구현하였다. 이 프로그램의 코드는 손의 움직임을 감지하여 마우스를 제어하는 기능을 구현한다. 이를 위해 cvzone 라이브러리의 HandDetector 클래스를 활용하며, 이 클래스의 인스턴스인 detector는 웹캠을 통해 손과 손가락의 위치를 판별하였다. 인스턴스 생성 시에는 감지 정확도와 동시에 감지할 수 있는 손의 최대 수를 설정할 수 있다. 웹캠을 활성화하고 영상을 읽어오기

위해 cv2.VideoCapture를 사용하였으며, 웹캠으로부터 한 프레임씩 이미지를 읽어와 img 변수에 저장한다. 이미지 내에서 손을 찾기 위해 detector.findHands 메서드를 사용하며, 이 메서드는 손의 위치 정보와 수정된 이미지를 함께 반환한다. 감지된 손의 손가락이 펼쳐져 있는지 아닌지를 판별하기 위해 detector.fingersUp 메서드를 사용하고 있다.¹³⁸⁾

이 메서드는 각 손가락의 상태를 나타내는 리스트를 반환하며, 손가락이 펼쳐져 있을 경우 1, 굽어 있을 때 0을 반환한다. 다만, 엄지의 경우 펼쳐져 있을 때 0, 아니면 1을 반환한다. 감지된 손의 개수가 2개인 경우에도 마찬가지로 각 손의 손가락 상태를 판별하며, 이때는 두 손의 정보를 각각 리스트로 반환한다. 이렇게 얻은 정보를 바탕으로 마우스의 동작을 결정하며, 마우스의 동작은 다양하게 구현될 수 있다.

손동작 상태 목록 리스트		
왼쪽	 [0, 1, 1, 1, 1]	 [1, 1, 1, 1, 1]
	 [1, 0, 0, 1, 1]	 [1, 0, 0, 0, 0]
오른쪽	 [0, 1, 1, 1, 1]	 [1, 1, 1, 1, 1]
	 [1, 0, 1, 1, 1]	 [1, 0, 0, 1, 1]

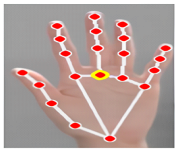
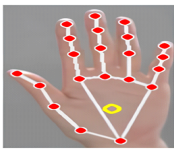
[그림3-7] 손동작 상태 리스트

138) <https://blog.learningbix.com/creating-ai-virtual-mouse/>

예를 들어, 모든 손가락이 펼쳐 마우스를 움직이며, 두 손을 특정한 동작으로 움직였을 때는 더블클릭하는 등의 동작을 구현하고 있다. 이런 방식으로, 이 코드는 사용자의 손동작을 감지하고 이를 마우스 제어에 활용함으로써 사용자에게 더 직관적인 인터페이스를 제공한다. 다음 [그림 3-7]은 손가락 상태를 `detector.fingersUp` 메서드로 변환한 리스트이다.

[그림 3-7]과 같이 반환된 손동작의 메서드 값을 감지하여 컴퓨터가 손동작을 인식하여 비디오월 제어 모니터의 UI의 아이콘 명령을 제어한다.

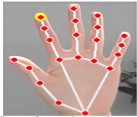
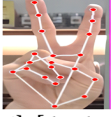
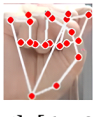
첫 번째로 마우스 이동에 대한 3가지 손동작[그림3-4]을 구분하여 가장 적합한 손동작 인식하도록 설계하였다. [그림 3-8]에서 손동작 1([1,1,0,0,0])이 감지되면, [그림 3-1]의 landmark 8번의 x, y 좌표 위치를 화면 좌표로 변환한 후, 이를 사용하여 마우스를 움직인다. 손동작 2([0, 1, 1, 1, 1])가 감지되면, [그림 3-1]의 landmark 9번의 x 좌표와 y 좌표 위치를 화면 좌표로 변환한 후, 이를 사용하여 마우스를 움직인다. 손동작 3([0, 1, 1, 1, 1])이 감지되면, [그림 3-1]의 landmark 9번과 0번의 중심인 x 좌표와 y 좌표 위치를 화면 좌표로 변환한 후, 이를 사용하여 마우스의 포인트를 움직여 컴퓨터의 제어화면으로 이동한다.

손동작 1	손동작 2	손동작 3
		
[1, 1, 0, 0, 0]	[0, 1, 1, 1, 1]	[0, 1, 1, 1, 1]

[그림 3-8] 마우스 이동 손동작

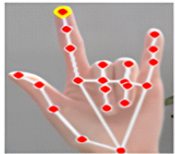
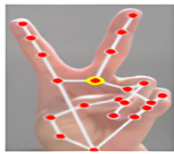
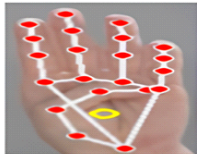
두 번째로 마우스 더블클릭을 하는 기능을 손동작 인식으로 전환하기 위해 [그림 3-9]와 같이 손동작 인식을 3가지로 구분하여 구현하였다. 손동작 1([0, 1, 1, 1, 1])이 감지되면, [그림 3-1]의 landmark 8번의 x, y좌표가 지정한 지점을 더블클릭한다. 손동작 2는 좌([1, 1, 1, 0, 0]), 우([0, 1, 1, 1, 1]) 동시에 감지되면 [그림 3-1]의 landmark 8번의 x, y 좌표가 지정한 지점

을 더블클릭한다. 손동작 3는 좌([1, 0, 0, 0, 0]), 우([0, 1, 1, 1, 1]) 동시에 감지되면 [그림 3-1]의 landmark 9번과 0번의 중심인 x, y 좌표가 지정한 지점을 더블클릭한다.

손동작 1	손동작 2	손동작 3
 [0, 1, 1, 1, 1]	좌  좌 [1, 1, 1, 0, 0] 우  우 [0, 1, 1, 1, 1]	좌  좌 [1, 0, 0, 0, 0] 우  우 [0, 1, 1, 1, 1]

[그림 3-9] 마우스 더블클릭 손동작

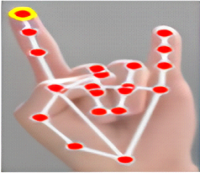
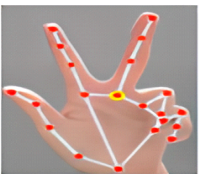
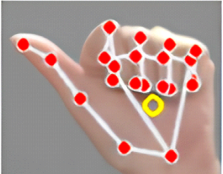
세 번째는 마우스 좌클릭에 대한 손동작 인식[그림 3-10]에서 손동작 1([0, 1, 0, 0, 1])이 감지되면, [그림 3-1]의 landmark 8번의 x, y 좌표가 가르키고 있는 지점을 좌클릭한다. 손동작 2([1, 1, 1, 0, 0])이 감지되면, [그림 3-1]의 landmark 9번의 x, y 좌표가 지정한 지점을 좌클릭한다. 손동작 3([1, 1, 1, 1, 1])이 감지되면, [그림 3-1]의 landmark 9번과 0번의 중심인 x, y 좌표가 지정한 지점을 좌클릭한다.

손동작 1	손동작 2	손동작 3
 [0, 1, 0, 0, 1]	 [1, 1, 1, 0, 0]	 [1, 1, 1, 1, 1]

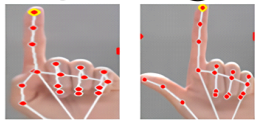
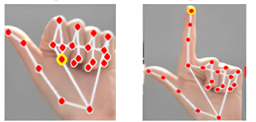
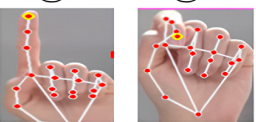
[그림 3-10] 마우스 좌클릭 손동작

네 번째는 마우스 우클릭에 대한 손동작 인식[그림 3-11]에서 손동작 1([1, 1, 0, 0, 1])이 감지되면, [그림 3-1]의 landmark 8번의 x, y 좌표가 가르키고 있는 지점을 우클릭한다. 손동작 2([0, 1, 1, 0, 0])이 감지되면, [그림

3-1]의 landmark 9번의 x, y 좌표가 지정한 지점을 우클릭한다. 손동작 3([0, 0, 0, 0, 0])이 감지되면, [그림 3-1]의 landmark 9번과 0번의 중심인 x, y 좌표가 지정한 지점을 우클릭한다.

손동작 1	손동작 2	손동작 3
		
[1, 1, 0, 0, 1]	[0, 1, 1, 0, 0]	[0, 0, 0, 0, 0]

[그림 3-11] 마우스 우클릭 손동작

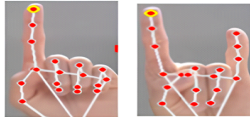
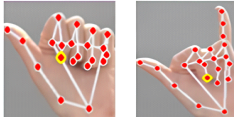
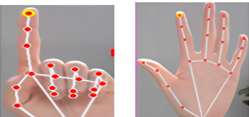
손동작 1	손동작 2	손동작 3
		
① [1, 1, 0, 0, 0] -> ② [0, 1, 0, 0, 0]	① [0, 0, 0, 0, 0] -> ② [0, 1, 0, 0, 0]	① [1, 1, 0, 0, 0] -> ② [1, 0, 0, 0, 0]

[그림 3-12] 마우스 스크롤다운 손동작

다섯 번째는 마우스 스크롤 다운에 대한 손동작 인식[그림 3-12]에서 손동작 1([1, 1, 0, 0, 0]에서 [0, 1, 0, 0, 0])로 순차적으로 감지되면, 마우스 휠을 아래로 일정한 단위만큼 이동한다. 손동작 2([0, 0, 0, 0, 0]에서 [0, 1, 0, 0, 0])로 순차적으로 감지되면, 마우스 휠을 아래로 일정한 단위만큼 이동한다. 손동작 3([1, 1, 0, 0, 0]에서 [1, 0, 0, 0, 0])로 순차적으로 감지되면, 마우스 휠을 아래로 일정한 단위만큼 이동한다.

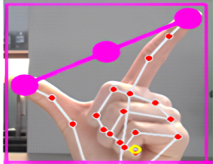
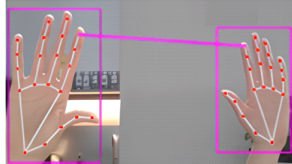
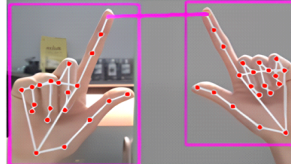
여섯 번째는 마우스 스크롤 업에 대한 손동작 인식[그림 3-13]에서 손동작 1([1, 1, 0, 0, 0]에서 [1, 1, 0, 0, 1])로 순차적으로 감지되면, 마우스 휠을 위로 일정한 단위만큼 이동한다. 손동작 2([0, 0, 0, 0, 0]에서

[0, 0, 0, 0, 1])로 순차적으로 감지되면, 마우스 휠을 위로 일정한 단위만큼 이동한다. 손동작 3([1, 1, 0, 0, 0]에서 [0, 1, 1, 1, 1])로 순차적으로 감지되면, 마우스 휠을 위로 일정한 단위만큼 이동한다.

손동작 1	손동작 2	손동작 3
 ① [1, 1, 0, 0, 0] -> ② [1, 1, 0, 0, 1]	 ① [0, 0, 0, 0, 0] -> ② [0, 0, 0, 0, 1]	 ① [1, 1, 0, 0, 0] -> ② [0, 1, 1, 1, 1]

[그림 3-13] 마우스 스크롤업 손동작

일곱 번째는 마우스 휠 업에 대한 손동작 인식[그림 3-14]에서 손모양 1([0, 1, 0, 0, 0])이 감지되면, [그림 3-1]의 landmark 8번과 4번의 x, y 좌표간 거리가 처음 웹 카메라에 나타났을 때의 거리를 기준으로 하여 기준보다 크면 거리에 비례하여 객체 크기를 확대한다.

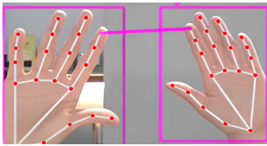
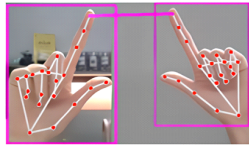
손동작 1	손동작 2	손동작 3
 [0, 1, 0, 0, 0]	 좌 [0, 1, 1, 1, 1] 우 [0, 1, 1, 1, 1]	 좌 [0, 1, 0, 0, 0] 우 [0, 1, 0, 0, 0]

[그림 3-14] 마우스 휠업 손동작

손 모양 2([0, 1, 1, 1, 1], [0, 1, 1, 1, 1])가 동시에 감지되면, 왼손과 오른손 각각의 [그림 3-1]의 landmark 8번의 x, y 좌표 간 거리를 측정한다. 웹캠에 처음 나타났을 때의 거리를 기준으로 하여 기준보다 크면 거리에 비

례하여 객체 크기를 확대한다. 손 모양 3([0, 1, 0, 0, 0], [0, 1, 0, 0, 0])이 동시에 감지되면, 왼손과 오른손 각각의 [그림 3-1]의 landmark 8번의 x, y 좌표 간 거리를 측정한다. 웹캠에 처음 나타났을 때의 거리를 기준으로 하여 기준보다 크면 거리에 비례하여 객체 크기를 확대한다.

여덟 번째는 마우스 휠 다운에 대한 손동작 인식[그림 3-15]에서 손모양 1([0, 1, 0, 0, 0])이 동시에 감지되면, [그림 3-1]의 landmark 8번과 4번의 x, y 좌표간 거리가 처음 웹 카메라에 나타났을 때의 거리를 기준으로 하여 기준보다 작으면 거리에 비례하여 객체 크기를 축소한다. 손모양 2([0, 1, 1, 1, 1], [0, 1, 1, 1, 1])가 동시에 감지되면, 왼손과 오른손 각각의 [그림 3-1] landmark 8번의 x, y 좌표 간 거리를 측정한다. 웹 카메라에 처음 나타났을 때의 거리를 기준으로 하여 기준보다 작으면 거리에 비례하여 객체 크기를 축소한다. 손 모양 3([0, 1, 0, 0, 0], [0, 1, 0, 0, 0])이 동시에 감지되면, 왼손과 오른손 각각의 landmark 8번의 x, y 좌표 간 거리를 측정한다. 웹캠에 처음 나타났을 때의 거리를 기준으로 하여 기준보다 작으면 거리에 비례하여 객체 크기를 줄인다.

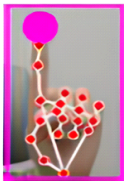
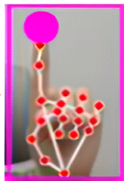
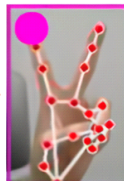
손동작 1	손동작 2	손동작 3
		
[0, 1, 0, 0, 0]	좌 [0, 1, 1, 1, 1] 우 [0, 1, 1, 1, 1]	좌 [0, 1, 0, 0, 0] 우 [0, 1, 0, 0, 0]

[그림 3-15] 마우스 휠다운 손동작

마지막으로 비디오월 제어시스템에서 이미지 및 동영상을 이동은 이동객체를 마우스로 클릭하여 도착지점까지 객체를 드래그로 이동하여 마우스 드래그를 푸는 손동작을 3가지 경우를 구성하여 가장 유효한 손동작이 어떤 손동작을 구현하였다.

[그림 3-16]의 손동작 1에서[1, 1, 0, 0, 0](=검지)이 감지되면, [그림 3-1]

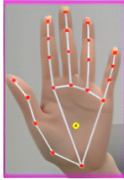
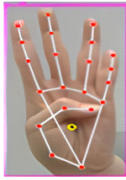
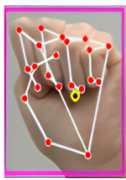
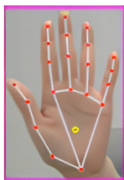
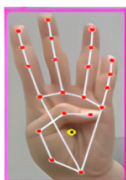
의 landmark 8번의 x 좌표와 y 좌표 위치를 화면 좌표로 변환한 후, 이를 사용하여 마우스를 움직인다. 이동시킬 객체 위치로 마우스 좌표가 놓이면, [1, 1, 1, 0, 0](=V자 손모양)이 감지되면, [그림 3-1] landmark 8번과 12번의 x, y 좌표 간의 거리가 30 정도 이하가 되면(30일 때는 검지와 중지가 거의 맞닿아 있는 경우) 마우스가 가리키고 있는 객체를 드래그하기 시작한다. [1, 1, 0, 0, 0](=검지)이 감지되면, [그림 3-1]의 landmark 8번의 x 좌표와 y 좌표 위치를 화면 좌표로 변환한 후, 이를 사용하여 마우스를 움직인다. 최종적으로 이동시킬 위치에 객체가 도달하면, [1, 1, 1, 0, 0](=V자 손 모양)이 감지시켜, [그림 3-1] landmark 8번과 12번의 x, y 좌표 간의 거리가 40 정도 이상이 되면(40일 때는 검지와 중지가 맞닿지 않고 V자가 되는 경우) 드래그를 풀고 객체를 내려놓는다.

손동작 1				
마우스 포인트 이동 손동작	이미지 및 영상 아이콘 선택 드래그 시작 손동작		마우스 포인트 이동 손동작	마우스 드래그 정지 손동작
				
[1, 1, 0, 0, 0]	[1, 1, 1, 0, 0]	[1, 1, 0, 0, 0]	[1, 1, 0, 0, 0]	[1, 1, 1, 0, 0]

[그림 3-16] 마우스 포인트 이미지 및 영상 이동 손동작 1

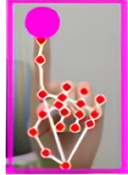
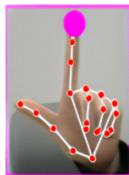
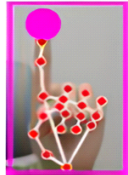
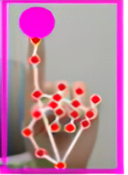
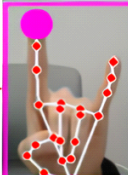
손동작 2의 [그림 3-17]에서 [0, 1, 1, 1, 1](=손가락 전체를 편다.)이 감지되면, [그림 3-1]의 landmark 9번의 x 좌표와 y 좌표 위치를 화면 좌표로 변환한 후, 이를 사용하여 마우스를 움직인다. 이동시킬 객체 위치로 마우스 좌표가 정해지면, [1, 1, 1, 1, 1](=엄지만 구부린다.) 손동작에서 [1, 0, 0, 0, 0](=주먹을 쥔다.) 손동작으로의 변화가 감지되면, 마우스가 가리키고 있는 객체를 드래그하기 시작한다. [0, 1, 1, 1, 1](=손가락 전체를 편다.)가 감지되면, [그림 3-1]의 landmark 9번의 x 좌표와 y 좌표 위치를 화면 좌표로 변

환한 후, 이를 사용하여 마우스의 포인트를 움직인다. 최종적으로 이동시킬 위치에 객체가 도달하면, [1, 1, 1, 1, 1](=엄지만 구부린다.)가 손동작을 감지시켜, 드래그를 풀고 객체를 내려놓는다.

손동작 2				
마우스 포인트 이동 손동작	이미지 및 영상 아이콘 선택 드래그 시작 손동작		마우스 포인트 이동 손동작	마우스 드래그 정지 손동작
				
[0, 1, 1, 1, 1]	[1, 1, 1, 1, 1]	[1, 0, 0, 0, 0]	[0, 1, 1, 1, 1]	[1, 1, 1, 1, 1]

[그림 3-17] 마우스 포인트 이미지 및 영상 이동 손동작 2

[그림 3-18]과 같이 손모양 3에서는 [1, 1, 0, 0, 0](=검지)이 감지되면, [그림 3-1]의 landmark 8번의 x 좌표와 y 좌표 위치를 화면 좌표로 변환한 후, 이를 사용하여 마우스의 포인트를 움직인다. 이동시킬 객체 위치로 마우스 좌표가 지정하면, [0, 1, 0, 0, 0](=검지)을 든 상태에서 엄지손가락을 들어 올림이 감지되면, 객체를 드래그하기 시작한다. [1, 1, 0, 0, 0](=검지)이 감지되면, [그림 3-1]의 landmark 8번의 x 좌표와 y 좌표 위치를 화면 좌표로 변환한 후, 이를 사용하여 마우스의 포인트를 움직인다. 최종적으로 이동시킬 위치에 객체가 도달하면, [1, 1, 0, 0, 1](=검지와 새끼손가락을 올린 손 모양)이 감지시켜, 드래그를 풀고 객체를 내려놓는다

손동작 3				
마우스 포인트 이동 손동작	이미지 및 영상 아이콘 선택 드래그 시작 손동작		마우스 커스 이동 손동작	마우스 드래그 정지 손동작
				
[1, 1, 0, 0, 0]	[0, 1, 0, 0, 0]	[1, 1, 0, 0, 0]	[1, 1, 0, 0, 0]	[1, 1, 1, 0, 1]

[그림 3-18] 마우스 포인트 이미지 및 영상 이동 손동작 3

위와 같이 컴퓨터 제어 모니터에서 제어되는 마우스 이동, 마우스 더블 클릭, 마우스 좌클릭, 우클릭, 마우스 스크롤다운/업, 마우스 휠업/다운, 이미지 및 영상을 이동하는 손동작을 보편적으로 자주 사용하는 각각 3가지 경우를 구성하여 인식률이 가장 높고 유용성이 가장 높은 손동작의 기능으로 구성하였다.

3) 음성 인식 기반 상호작용 설계

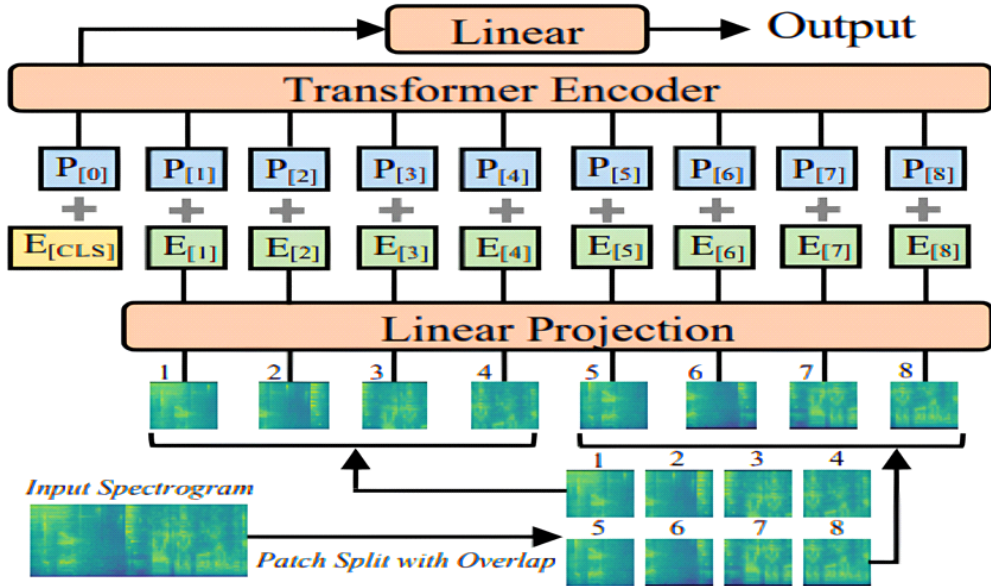
가) 설계 방법

음성 인식을 구현하기 위해 "MIT/ast-finetuned-speech-commands-v2"[그림3-19]는 MIT에서 개발한 음성 인식 모델이다.¹³⁹⁾ 이 모델은 30개 이상의 단일 단어 음성 명령으로 이루어진 Google의 Speech Commands 데이터 셋을 사용하여 사전 학습된 후, 추가적인 미세 조정이 되었다.¹⁴⁰⁾

이 모델을 활용하여 사용자의 음성 명령을 텍스트로 변환하는 시스템을 구현하였다. 사용자가 특정 단어를 발음하면, 이 음성은 텍스트로 전환되어 'voice_label.txt' 파일에 저장된다. 이 파일은 음성 인식 코드와 똑같은

139) Gong, Y., Chung, Y. A., & Glass, J. (2021). Ast: Audio spectrogram transformer. arXiv preprint arXiv:2104.01778.

140) <https://huggingface.co/MIT/ast-finetuned-speech-commands-v2>



[그림 3-19] 오디오 스펙트럼 트랜스포머 아키텍처

폴더에 위치한다. 저장된 텍스트는 다시 읽어와서 그에 해당하는 기능을 실행시키는 데 사용된다. 예를 들어, 사용자가 ‘left’, ‘right’, ‘two’, ‘up’, ‘down’ 등의 단어를 발음하면, 이에 부합하는 특정 기능이 실행된다. 이렇게 하면 사용자는 음성 명령만으로도 다양한 기능을 제어할 수 있게 된다.

먼저 마우스 움직임을 음성 인식으로 제어하는 것은 위치 정보가 필요하고 음성 인식만으로는 어려움이 있어 사용자가 원하는 마우스의 포인트를 정확하게 지정할 필요가 있다. 음성으로만 이를 표현하는 것은 제한적이기 때문에 본 연구에서는 다양한 손 모양을 활용하여 마우스의 포인트 이동을 구현하였다. 마우스 포인트 이동[그림 3-4]의 세 가지 손동작을 활용하여 검증하고자 한다. 이 세 가지 손동작 중, 손동작 2와 손동작 3는 마우스 위치에 대응하는 포인트가 상대적으로 아래에 있어, 사용자가 손을 더 많이 높여야 하는 단점이 있다. 이러한 사용자의 피드백을 반영하여, 사용자가 가장 편리하게 사용할 수 있는 마우스 이동 손동작 1을 활용하여 마우스 위치 정보를 화면에 표시하도록 구현하였다. 이것은 사용자는 손을 크게 움직이지 않아도 마우스를 원하는 방향과 속도로 움직일 수 있다.

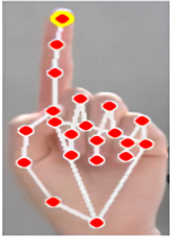
우선 사용자가 제어화면의 제어 아이콘을 마우스 더블클릭을 하기 위해

손동작	음성 명령	제어화면 마우스 동작
 [1, 1, 0, 0, 0]		마우스 더블 클릭
		마우스 좌클릭
		마우스 우클릭

[그림 3-20] 음성 인식 활용한 화면제어1

음성 인식으로 적용하려면 [그림 3-20]에서 처럼 손동작([1, 1, 0, 0, 0])이 감지되면, 손가락 끝 지점의 좌표인 [그림 3-1]의 landmarks의 8번 좌표 (x, y)를 화면 좌표로 변환하여 마우스를 움직인 다음 사용자가 ‘Two’ 단어를 발음하면, 이 음성이 텍스트로 전환되어 ‘voice_label.txt’ 파일에 ‘Two’라는 단어로 저장된다. 그리고 이 파일을 읽어 ‘Two’ 단어가 저장된 것을 확인 후 현재 마우스가 위치한 지점을 더블클릭하도록 프로그램이 작동한다. 이렇게 하면 사용자는 손동작과 음성 명령을 통해 마우스의 더블 클릭 기능을 제어할 수 있게 된다. 그리고 사용자가 마우스 좌클릭 및 우클릭을 제어하려면 먼저 마우스를 원하는 지점에 있어야 하고, 이 동작을 수행하기 위해서는 사용자의 손동작을 인식하여 마우스 포인트를 이동한 후 음성 인식을 수행하여 마우스 좌클릭 및 우클릭을 구현하였다. [그림 3-8] 손동작([1, 1, 0, 0, 0])이 감지되면, 손가락 끝의 좌표인 [그림 3-1]의 landmarks의 8번 좌표 (x, y)를 화면의 좌표로 변환하여 마우스를 움직인 후, 사용자가 ‘Left’ 또는 ‘Right’라는 단어를 발음하면, 이 음성이 텍스트로 전환되어 ‘voice_label.txt’ 파일에 ‘Left’ 또는 ‘Right’라는 단어로 저장된다. 그리고 이 파일을 읽어 ‘Left’ 또는 ‘Right’라는 단어가 저장되어 있을 때, 현재 마우스가 위치한 지점

을 좌클릭 또는 우클릭을 하도록 프로그램이 작동하여 사용자의 손동작과 음성 명령을 통해 마우스의 좌클릭 및 우클릭 기능을 제어할 수 있게 된다.

손동작	음성 명령	제어화면 마우스 동작
 [1, 1, 0, 0, 0]		마우스 스크롤 다운 (스크롤 위치 지정)
		마우스 스크롤 업 (스크롤 위치 지정)
		마우스 휠 업(확대) (특정 객체 지정)
		마우스 휠 다운(축소) (특정 객체 지정)

[그림 3-21] 음성 인식 활용한 화면제어2

세 번째로는 제어화면을 마우스 스크롤다운 및 업을 기능을 음성으로 제어하려면 [그림 3-21] 와 같이 손동작([1, 1, 0, 0, 0])으로 마우스 포인트를 이동하여 원하는 지점에 있고 손동작을 인식과 동시에 음성으로 명령을 수행한다. 손모양([1, 1, 0, 0, 0])이 감지되면, 손가락 끝의 좌표인 [그림 3-1]의 landmarks의 8번 좌표 (x, y)를 화면의 좌표로 변환하여 마우스를 움직인 후, 사용자가 ‘Down’ 또는 ‘Up’이라는 단어를 발음하면, 이 음성이 텍스트로 전환되어 ‘voice_label.txt’ 파일에 ‘Down’ 또는 ‘Up’이라는 단어로 저장된다. 그리고 이 파일을 읽어 ‘Down’ 또는 ‘Up’이라는 단어가 저장된 것을 확인 후, 현재 마우스가 위치한 지점을 스크롤다운 및 업이 되도록 프로그램이 작동하여 사용자는 손동작과 음성 명령을 통해 마우스의 스크롤다운 및 업 기능을 컴퓨터의 화면에서 제어할 수 있게 된다.

마지막으로 마우스 휠업 및 다운에 대한 기능을 음성으로 수행하려면 [그

림 3-21] 와 같이 손동작([1, 1, 0, 0, 0])을 특정 이미지나 영상을 지정한 상태에서 ‘Up’ 또는 ‘Down’의 음성을 인식하여 명령을 수행한다. [그림 3-21]의 손동작 모양([1, 1, 0, 0, 0])이 감지되면, 손가락 끝의 좌표인 [그림 3-1]의 landmarks의 8번 좌표 (x, y)를 화면의 좌표로 변환하여 마우스를 움직인 후, 사용자가 ‘Up’ 또는 ‘Down’이라는 단어를 발음하면, 이 음성이 텍스트로 전환되어 ‘voice_label.txt’ 파일에 ‘Up’ 또는 ‘Down’이라는 단어로 저장된다. 그리고 이 파일을 읽어 ‘Up’ 또는 ‘Down’이라는 단어가 저장된 것을 확인 후, 현재 마우스가 위치한 지점을 휠 업 및 다운하여 객체의 크기를 확대 및 축소하도록 프로그램이 작동하여, 손동작과 음성 명령을 통해 마우스 휠업(확대) 및 다운(축소) 기능을 제어화면을 제어할 수 있다.

이처럼 음성 인식만으로 마우스를 이동이나 더블클릭, 좌클릭, 우클릭, 스크롤다운, 스크롤업, 휠업/다운은 수행하기에는 어려움이 있어 손동작과 동시에 음성을 인식하는 방법을 사용하여 컴퓨터 제어화면을 제어하였고, 이미지나 영상의 객체를 이동하여 화면 확대, 축소 기능의 명령을 GUI 상에서 음성으로 정확하게 나타내지 못하여 손동작과 결합하여 구현하였다.

4) 손동작과 음성 인식 기반 비디오월 제어시스템 상호작용 설계

가) 구현 방법

본 연구에서 비디오월 제어의 상호작용 구현에서 손동작과 음성 인식을 동시에 구현하여 활용성과 유용성을 검증하기 위해 손동작 인식은 구글 사의 머신 러닝 기반 고성능 손 제스처 인식 솔루션인 MediaPipe Hands¹⁴¹⁾와 음성 인식인 ‘MIT/ast-finetuned-speech-commands-v2’의 모델¹⁴²⁾을 결합하여 비디오월의 제어시스템을 제어하는 구현 방법 모델을 적용하였다. 특정 손동작과 음성 명령을 동시에 기능이 수행되어 제어시스템의 인식률과 정확성을 향상을 높이는 기능을 수행할 수 있도록 설계하였다.

먼저 비디오월 제어시스템의 제어 모니터에서 마우스 더블클릭을 활용하여 제어 아이콘을 구현하기 위해 설계로 [그림 3-22]에서 손동작([0, 1, 1, 1,

141) <https://developerai.com/10>

142) <https://huggingface.co/MIT/ast-finetuned-speech-commands-v2>

1))으로 손의 포인트인 [그림 3-1]의 landmark 8번의 x, y 좌표를 화면의 위치 정보를 가져와서 그 위치에서 마우스 포인트를 제어한다. 다음 단계에서는 사용자의 음성 명령을 수행한다.

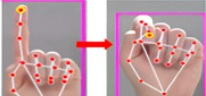
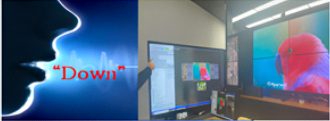
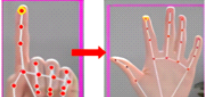
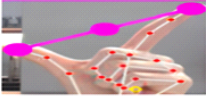
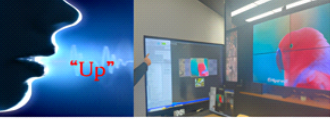
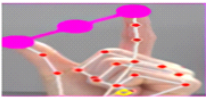
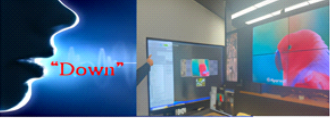
손동작	음성 명령	제어화면 마우스 동작
 [0, 1, 1, 1, 1]	 "Two"	마우스 더블 클릭
 [0, 1, 0, 0, 1]	 "Left"	마우스 좌클릭
 [1, 1, 0, 0, 1]	 "Right"	마우스 우클릭

[그림 3-22] 손동작과 음성인식 제어시스템 1

사용자가 'Two'로 명령하면, 음성이 텍스트로 변환되어 'voice_label.txt' 파일에 'Two'라는 단어로 저장된다. 이 파일을 읽어 'Two'라는 단어가 저장되어 있어 현재 마우스가 위치한 지점에서 더블클릭하는 동작을 실행하고, 사용자는 손동작과 음성 명령을 결합하여 마우스의 더블클릭 기능을 제어할 수 있다.

두 번째는 제어시스템의 이미지나 영상을 마우스로 좌클릭이나 우클릭 명령을 수행하는 구현 방법을 설계하였다. [그림 3-22]에서 마우스 좌클릭 손동작([0, 1, 0, 0, 1])과 우클릭 손동작([1, 1, 0, 0, 1])을 나타내면 손의 포인트가 [그림 3-1]의 landmark 8번의 x, y 좌표를 위치 정보를 가져와서 마우스 움직임을 감지한 다음에 사용자의 음성 인식하여 명령을 수행한다. 사용자가 좌클릭 손동작과 'Left'라고 발음하거나, 우클릭 손동작과 'Right'라고 발음을 하면, 각 음성이 텍스트로 변환되어

‘voice_label.txt’ 파일에 ‘Left’ 또는 ‘Right’라는 단어로 저장되어, 이 파일을 읽어 ‘Left’ 또는 ‘Right’라는 단어가 저장되어 있을 것을 확인 후, 현재 마우스가 위치한 지점에서 좌클릭이나 우클릭 동작을 실행한다. 이렇게 사용자는 손동작과 음성 명령을 결합하여 비디오월 제어시스템의 마우스의 좌/우클릭 기능을 제어하도록 하였다.

손동작	음성 명령	제어화면 마우스 동작
 [1, 1, 0, 0, 0] [1, 0, 0, 0, 0]	 "Down"	마우스 스크롤 다운 (스크롤 위치 지정)
 [1, 1, 0, 0, 0] [0, 1, 1, 1, 1]	 "Up"	마우스 스크롤 업 (스크롤 위치 지정)
 [0, 1, 0, 0, 0]	 "Up"	마우스 휠 업(확대) (특정 객체 지정)
 [0, 1, 0, 0, 0]	 "Down"	마우스 휠 다운(축소) (특정 객체 지정)

[그림 3-23] 손동작과 음성 인식 제어시스템 2

마지막으로 비디오월 제어화면에서 마우스 스크롤다운/업과 마우스 휠업/다운을 제어하는 기능을 설계하였다. [그림 3-23]과 같이 사용자가 마우스 스크롤 다운을 나타내는 손동작([1, 1, 0, 0, 0] → [1, 0, 0, 0, 0])과 마우스 스크롤 업을 나타내는 손동작([1, 1, 0, 0, 0] → [0, 1, 1, 1, 1])을 웹캠에 인식시키면, 이 손의 포인트인 [그림 3-1]의 landmark 8번의 x, y 좌표를 화면의 위치 정보로 가져와서 그 위치는 마우스의 움직임을 제어하며, 마우스 스크롤다운의 손동작과 ‘Down’의 음성을 인식, 마우스 스크롤업의 손동작과 ‘Up’이라는 음성을 인식시키면 이 음성이 텍스트로 변환되어 ‘voice_label.txt’ 파일에 ‘Down’ 또는 ‘Up’이라는 단어로 저장되

며, 이 파일을 읽어 ‘Down’ 또는 ‘Up’이라는 단어가 저장되어, 현재 마우스가 위치한 지점에서 스크롤업 또는 다운하는 동작을 실행하여 사용자는 손동작과 음성 명령을 결합하여 마우스 스크롤업 및 다운 기능을 제어할 수 있다.

그리고 마우스 휠 업 및 다운 기능을 제어하기 위해서는 [그림3-23]와 같이 마우스 휠업 손동작([0, 1, 0, 0, 0])과 마우스 휠다운 손동작([0, 1, 0, 0, 0])을 웹캠에 인식시켜서 [그림 3-1] landmark 8번과 4번의 x, y 좌표 간의 거리를 측정한다. 이 측정된 거리는 웹캠에 처음 손이 나타났을 때의 거리 차를 기준으로 작거나 크게 사용자가 ‘Up’ 이나 ‘Down’이라고 음성을 입력하면 측정된 거리에 비례하여 지정 이미지나 영상의 객체의 크기를 축소하거나 확대하는 동작을 실행하여 사용자는 비디오월의 제어화면을 손동작과 음성 명령을 결합하여 마우스의 스크롤 휠다운(축소), 업(확대) 기능을 제어할 수 있다.

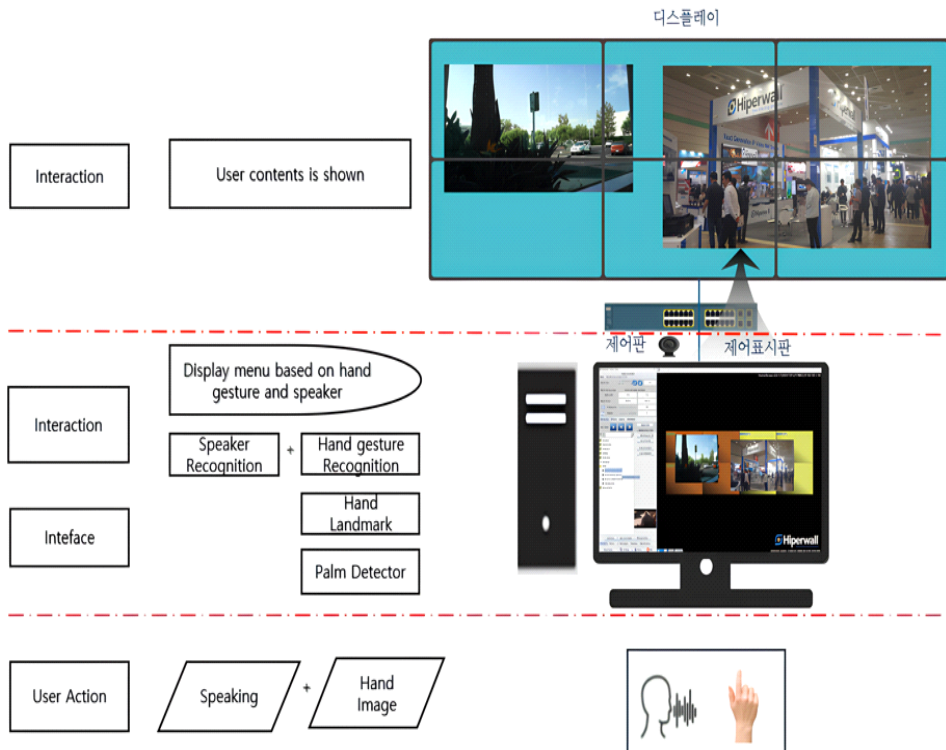
비디오월의 제어화면에서 마우스로 제어화면의 이미지나 영상을 자유자재로 이동시키는 작업에서 음성 인식을 적용하는 것은 두 가지 어려움이 존재하는데 첫째는 이동할 목적지를 음성으로 정확하게 지시하는 부분, 둘째는 마우스의 드래그를 통해 이미지나 영상의 객체를 이동시킬 때 필요한 정확한 궤적을 음성으로 표현하는 것은 상당한 기술을 필요하므로 본 연구에서는 마우스로 이미지나 영상을 이동시키는 작업을 손동작만으로 제어하도록 설계하였다.

제 4 장 구현 및 테스트

제 1 절 구현 범위

1) 제어시스템 범위

제3장에서 설계된 멀티모달 기반 비디오월 제어시스템의 상호작용에 대한 구현 알고리즘은 [그림 4-1]과 같이 손동작과 음성을 인식하여 비디오월의 제어시스템을 제어하기 위해 구현하였고 비디오월 제어화면은 H사의 비디오월 제어화면을 활용하였고 H사의 비디오월 제어화면의 기능과 손동작과 음성 인식으로 제어하는 인터페이스 기능 설명은 [표 4-1]과 같다.



[그림 4-1] 손동작 및 음성 인식 비디오월 제어 구조

[표 4-1] 손동작 및 음성 인식 비디오월 제어별 기능

구분	단 계	동작 및 사양
User Action	Action	· 사용자 손동작 이미지 · 사용자 음성 명령어
Interface	Hand Recognition	· Divice : 웹 카메라 · 개발언어 : MediaPipe, OpenCV, Python(3.8.18)
	Speaker Recognition	· Divice : 스피커 · 개발언어 : Python(3.11.5)
	Application	· hiperwall contoller v8.5
	Hardware	· 13세대 인텔® 코어™ i9-13900K (32 MB 캐시, 24 코어, 3.00 GHz · 32 GB: 2 x 16 GB · NVIDIA® GeForce RTX™ 4080(16G) · HDD 1TB
Interaction	제어판 폴더지정 제어	· 손동작 및 음성 인식으로 제어
	제어판 스크롤 바 제어	· 손동작 및 음성 인식으로 제어
	제어표시판 콘텐츠 삭제	· 손동작으로 콘텐츠 지정 후 음성 명령
	제어표시판 콘텐츠 확대	· 손동작으로 콘텐츠 지정 후 음성 명령
	제어표시판 콘텐츠 축소	· 손동작으로 콘텐츠 지정 후 음성 명령

2) 제어시스템 인프라 구성

가) 하드웨어 구성

비디오월 제어시스템에 손동작 및 음성 인식을 제어하는 연구 수행을 위하여 하드웨어 구성은 [표 4-2]와 같이 Microsoft의 윈도우 OS 환경에서 13세대 인텔 CPU를 탑재한 데스크탑에 웹 카메라와 스피커를 장착하여 비디오월 제어 하드웨어 시스템을 구성하였고, 응용 소프트웨어인 hiperwall 8.5와 개발 언어로 파이썬, OpenCV를 설치하여 환경 구축을 완료하였다. 또한 비디오월 제어시스템과 상호작용을 하기 위해 50" 모니터를 8개를 연결하여 대형 모니

터를 구성하여 비디오월 제어시스템에서 손동작 및 음성 인식으로 제어가 되면 8개 모니터에 바로 표출되도록 환경을 구성하였다.

나) 개발 환경 및 소프트웨어 구성

[표 4-2]는 손동작 인식은 구글이 개발한 오픈소스 라이브러리인 MediaPipe Hands를 활용하여 손을 감지하고 추적하는 기술을 응용하였고, 개발언어는 파이썬으로 개발하였고, 라이브러리들인 OpenCV, cvzone의 HandDetector, mouse, numpy 등이 import하였고, cvzone의 HandDetector 객체는 웹캠 이미지에서 손을 찾는 데 사용되었다. 손이 감지되면, 해당 손의 동작에 맞는 기능을 실행하도록 하였다. 그리고 음성 인식은 스피커를 통해 음성 명령을 내리면 비디오월 제어시스템에서 명령이 수행하도록 하였다.

[표 4-2] 손동작 및 음성 인식을 위한 하드웨어 및 소프트웨어 구조

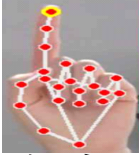
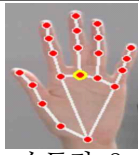
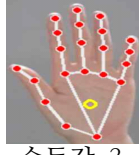
구분	내용
손동작 인식	• Python, OpenCV • windows 10 • 비전 AI 라이브러리 : MediaPipe • Device : 웹 카메라 • Hardware : Desktop PC(i9-13900K, 32G, 1T) NVIDIA® GeForce RTX™ 4080(16G)
음성 인식	• Python • windows 10 • 음성인식 모델 : MIT/ast-finetuned-speech-commands-v2 • Device : 스피커 • Hardware : Desktop PC(i9-13900K, 32G, 1T) NVIDIA® GeForce RTX™ 4080(16G)
Interface	• Application : hiperwall 8.5 • Hardware : Desktop PC(i9-13900K, 32G, 1T) NVIDIA® GeForce RTX™ 4080(16G)
Interaction	• windows 10 • 50" LCD Monitor • Monitor display Decoder: i7 9700T/SSD128G/8G

제 2 절 시스템 성능 구현 및 사용성 평가

1) 성능 구현 및 평가

비디오월 제어시스템에서 손동작으로 마우스 이동, 마우스 좌/우클릭, 마우스 더블클릭, 마우스 스크롤업/다운, 마우스 휠업(화면 확대)/다운(화면 축소) 기능을 3가지 손동작으로 구현하여 테스트한 결과 [표 4-3]에서 나타내주고 있다.

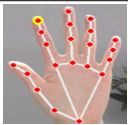
[표 4-3] 3가지 손동작 구현 테스트 결과 1

손동작	테스트 참여자									제어판 기존 제어 방식
	A		B		C		계			
	횟수	성공 횟수	횟수	성공 횟수	횟수	성공 회수	횟수 계	성공 횟수 계	성공률	
 손동작 1	20	20	20	20	20	20	60	60	100%	마우스 포인트 이동
 손동작 2	20	20	20	20	20	20	60	60	100%	
 손동작 3	20	20	20	20	20	20	60	60	100%	

비디오월 제어시스템의 제어판 UI 아이콘을 마우스 포인트를 이동하여 폴더를 지정할 때 음성 인식으로 제어하는 것보다 손동작 인식으로 제어하는 것이 더 편리한 것을 확인하였고 보편적으로 자주 사용하는 3가지 손동작으로 테스트를 해 본 결과 3가지 모두 100%의 인식률이 도출되었다. [그림

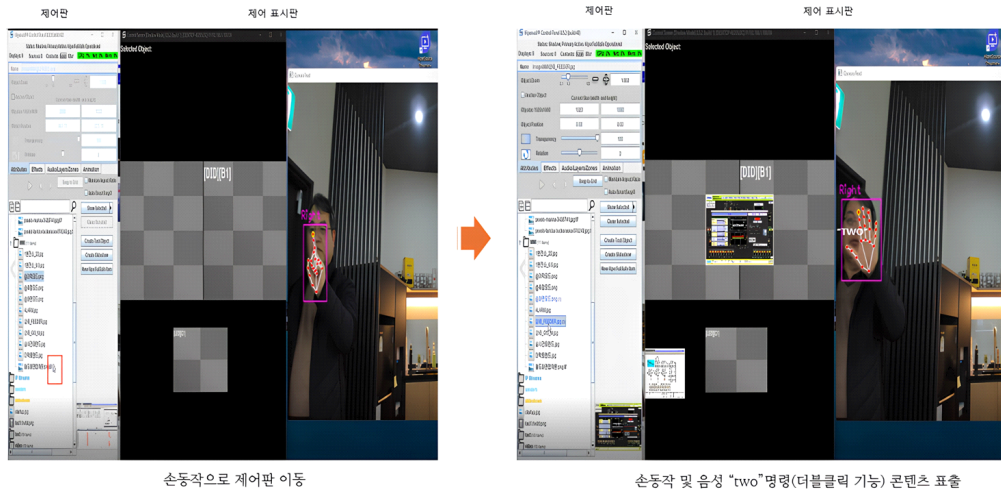
3-1]과 [표 4-3]에서 손동작 2의 landmark 9번과 손동작 3의 landmark 0과 9번 사이가 손동작 1의 landmark 8번보다 x, y좌표를 지정하는 기준점이 아래에 위치하여 손을 더 많이 올려서 포인트를 이동하여야 하므로 실험 참석자들이 다소 불편하였고, 선호도 측면에서 손동작 1번이 우수하여 반영하였다. 음성 인식으로 마우스를 이동하는 것은 기술적인 한계로 인해 차후 연구 과제로 진행하기로 하였고, 손동작 1번만으로 마우스를 이동하는 것을 채택하기로 했다. 다음은 마우스 더블클릭에 대한 3가지 손동작을 테스트해 본 결과 [표 4-4]의 결과를 얻었다.

[표 4-4] 3가지 손동작 구현 테스트 결과 2

손동작	테스트 참여자									제어판 기존 제어 방식
	A		B		C		계			
	횟수	성공 횟수	횟수	성공 횟수	횟수	성공 회수	횟수 계	성공 횟수 계	성공률	
 손동작 1	20	16	20	18	20	16	60	50	83.3%	마우스 더블 클릭
 손동작 2	20	16	20	15	20	17	60	48	80.0%	
 손동작 3	20	17	20	16	20	16	60	49	81.7%	

[표 4-4]와 같이 양쪽 손모양으로 마우스 더블 클릭 기능을 제어판 UI에서 폴더를 제어할 때 성공률 80%, 81.7%보다 한 손동작으로 제어하는 것이 성공률이 83.3%로 상대적으로 높게 나타났다. [표 4-3]의 손동작 1로 마우스를 포인트를 이동하여 검지를 제외한 손가락을 펼쳤을 때 마우스 더블클릭 기능으로 전환되어 제어판 UI 폴더에서 연속 두 손동작 즉 [표 4-3]의 손동작 1과 [표 4-4] 손동작 1의 동작으로 더블클릭 기능으로 구현할 수 있었다. 본

연구에서 더블클릭 기능을 음성 인식으로 추가하여 [표 4-3]의 손동작1을 사용하여 음성으로 “two” 하는 명령을 내리면 비디오월 제어시스템의 제어판의 UI 폴더가 더블클릭 기능을 구현되는 것을 확인하였다. [그림 4-2]는 구현 테스트를 진행한 화면이다.

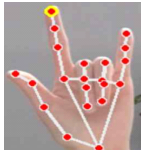
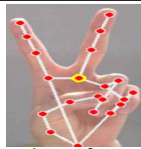
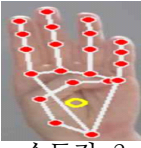
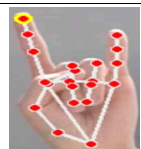
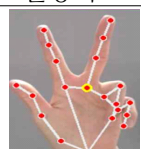
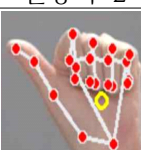


[그림 4-2] 손동작 및 음성 명령 콘텐츠 호출

제어판에서 이미지, 동영상, 웹브라우저, IP스트림 영상을 불러올 수 있는 폴더를 더블 더블 클릭하면 각각의 소스 콘텐츠들이 제어표시판에 표시되고 이 호출된 콘텐츠들은 TCP/IP와 연결된 출력 소프트웨어와 연동이 되어 있는 대형 디스플레이 화면에 표시가 되는 구조이다. [그림 4-2]는 손동작과 음성으로 기존의 마우스로 제어하는 폴더를 클릭과 더블클릭을 하여 콘텐츠를 호출하는 방식으로 전환하는 방식으로 테스트를 진행하였고 손동작과 음성으로 제어할 경우 성공률은 90%에 미치지 못한 결과를 얻었다. 마우스 이동하는 기능을 손동작으로 했을 때 보다 상대적으로 낮은 수치임을 알 수 있었다.

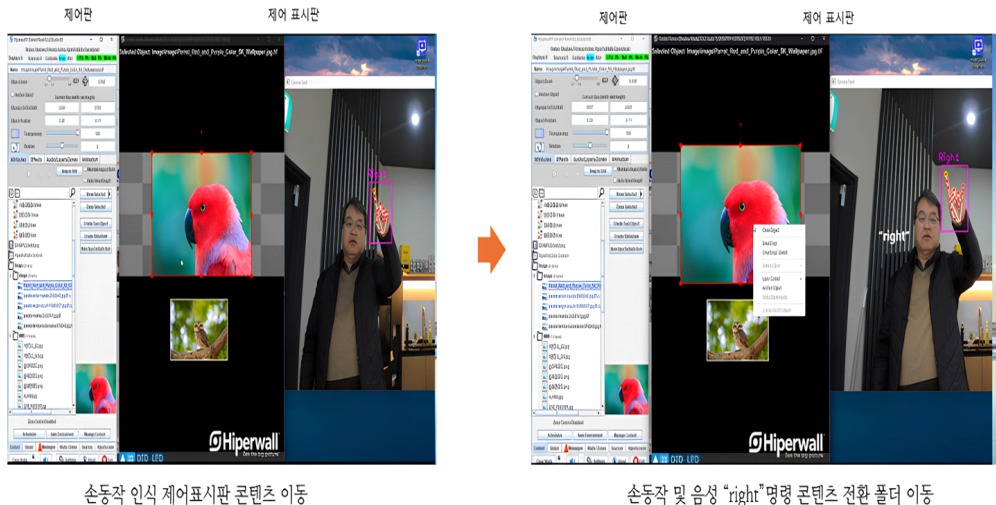
3장의 기능 정의 2로 마우스 좌클릭 우클릭 기능을 손동작 및 음성으로 제어하기 위해 3가지 손동작을 테스트하고 그 동작중 인식 성공률과 선호도가 높은 동작을 음성과 동시에 제어가 가능하도록 표시판의 콘텐츠를 제어하였다. {표 4-5}는 3가지 손동작에 대한 테스트 결과표이다.

[표 4-5] 3가지 손동작 구현 테스트 결과 3

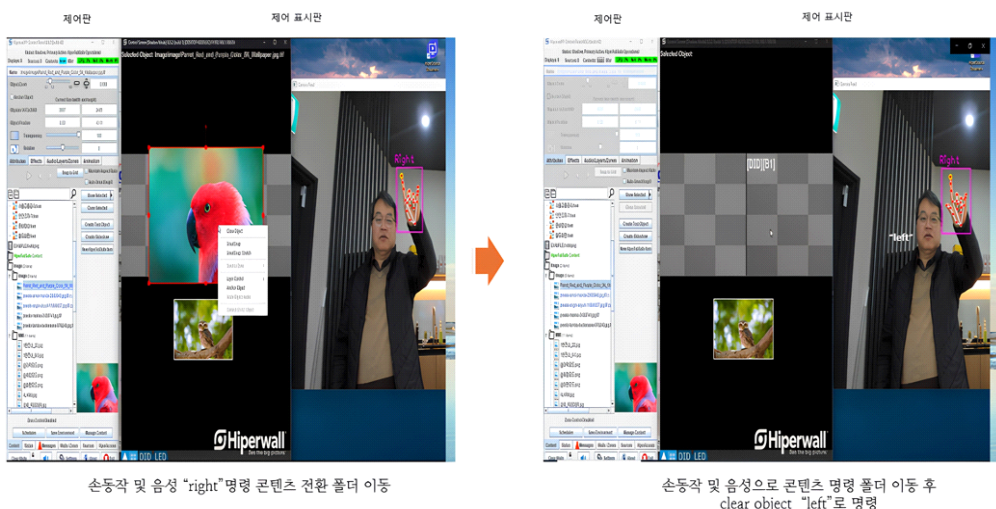
손동작	테스트 참여자									제어판 기존 제어 방식
	A		B		C		계			
	횃수	성공 횃수	횃수	성공 횃수	횃수	성공 회수	횃수 계	성공 횃수 계	성공률	
 손동작 1	20	18	20	19	20	19	60	56	93.3%	마우스 좌클릭
 손동작 2	20	17	20	16	20	16	60	49	81.7%	
 손동작 3	20	19	20	18	20	19	60	56	93.3%	
 손동작 1	20	16	20	17	20	18	60	51	85.0%	마우스 우클릭
 손동작 2	20	19	20	20	20	19	60	58	96.7%	
 손동작 3	20	20	20	20	20	20	60	60	100%	

[표 4-5]와 같이 마우스 좌클릭의 손동작으로 손동작 1은 93.3%, 손동작 2는 81.7%, 손동작 3은 93.3%로 제어에 대한 성공률을 나타냈으며 손동작 1과 손동작 3이 성공률이 높았다. 또한 마우스 우클릭의 손동작으로 손동작 1

은 85%, 손동작 2는 96.7%, 손동작 3는 100%의 성공률로 측정되었고, 손동작 3이 성공률이 가장 높았다. [그림 4-3]과 [그림 4-4]는 손동작과 음성으로 제어판에서 제어표시판에 콘텐츠가 표출되면 제어표시판이 콘텐츠를 마우스 좌클릭 및 우클릭으로 제어하는 기능을 손동작으로 콘텐츠를 지정하고 음성으로 'right', 명령을 내리면 콘텐츠의 다음 명령 폴더가 표시되고, 다시 손동작으로 표시된 폴더의 명령 아이콘 clear object 지정하고 'left'로 명령을 내리고 콘텐츠가 제거되는 테스트를 시연한 장면이다.



[그림 4-3] 손동작 콘텐츠 이동하여 음성 right 명령

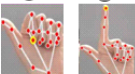
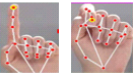
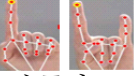
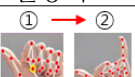
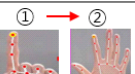


[그림 4-4] 손동작 콘텐츠 이동하여 음성으로 left 명령

다음은 기능 정의 3으로 마우스를 제어판에 스크롤업/다운하는 기능을 손동작 및 음성으로 제어하는 테스트를 3가지 손동작을 통해 인식 성공률을 측정하여 가장 선호도가 높은 손동작과 음성 명령으로 내렸을 때 제어가 가능하지 구현하였다. [표 4-6]는 마우스 스크롤 업/다운의 3가지 손동작으로 적용한 테스트 결과이다.

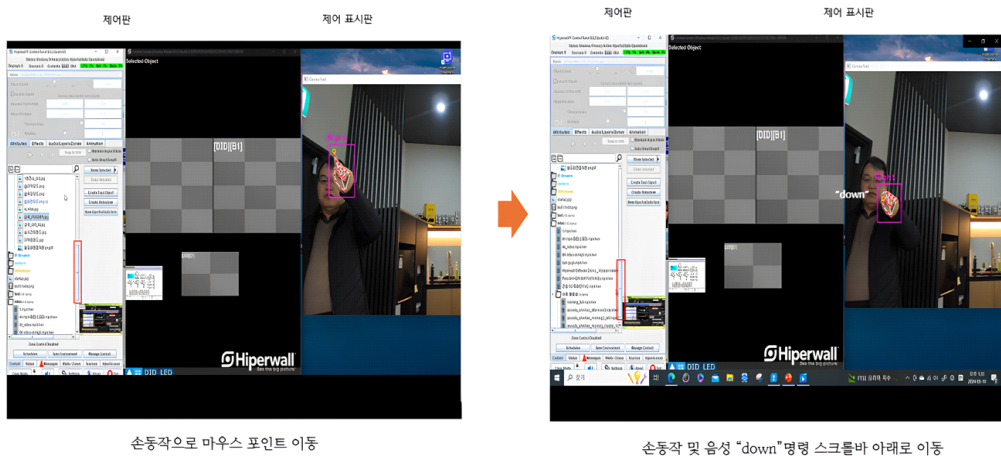
제어판의 기존 마우스로 스크롤바를 아래로 내리는 기능을 손동작으로 테스트를 구현하였는데 손동작 1과 2는 100% 성공률이 나타났고, 손동작 3은 91.7%의 성공률로 나타났다. 손동작 1은 사용자 입장에서 편하다는

[표 4-6] 3가지 손동작 구현 테스트 결과 4

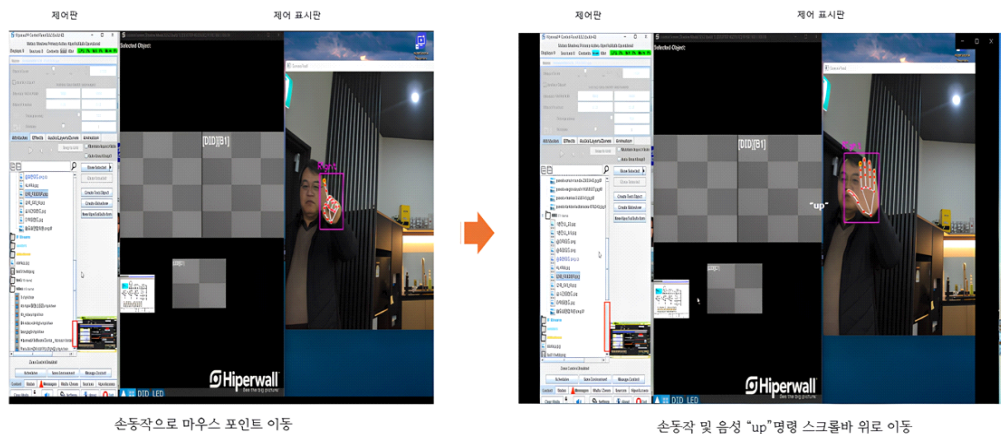
손동작	테스트 참여자									제어판 기존 제어 방식
	A		B		C		계			
	횏수	성공 횏수	횏수	성공 횏수	횏수	성공 회수	횏수 계	성공 횏수 계	성공 률	
① → ②  손동작 1	20	20	20	20	20	20	60	60	100%	마우스 스크롤 다운 (down)
① → ②  손동작 2	20	20	20	20	20	20	60	60	100%	
① → ②  손동작 3	20	18	20	19	20	18	60	55	91.7%	
① → ②  손동작 1	20	20	20	20	20	20	60	60	100%	마우스 스크롤 업 (up)
① → ②  손동작 2	20	20	20	20	20	20	60	60	100%	
① → ②  손동작 3	20	19	20	19	20	19	60	57	95%	

평가되었고, 손동작 2는 검지를 구부렸다 폈다하는 동작에서 다소 불편함이 평가되었다, 손동작 3은 검지만 남기고 나머지 손가락을 펴는 동작으로 손동작 2와 같이 불편했다.

손동작 3은 검지만 남기고 나머지 손가락을 펴는 동작으로 손동작 2와 같이 불편했다. 그리고 스크롤바를 위로 올리는 기능을 3가지 손동작으로 구현하였는데 손동작 1과 손동작 2는 100% 성공률이 나타났고, 손동작 3은 95%의 성공률로 나타났다. 손동작으로 스크롤바를 통해 인식 성공률을 측정하여 가장 선호도가 높은 손동작과 음성으로 스크롤바를 올리거나 내리는 명령 'up', 'down'을 내렸을 때 제어가 가능하도록 [그림 4-5]과 [그림 4-6]을 구현하고 테스트한 과정을 캡처한 장면이다.



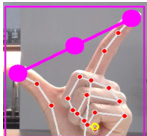
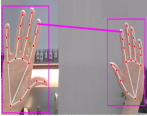
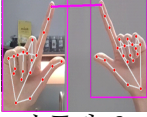
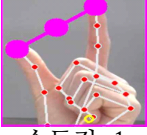
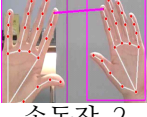
[그림 4-5] 손동작 및 음성명령 제어판 스크롤바 아래로 제어



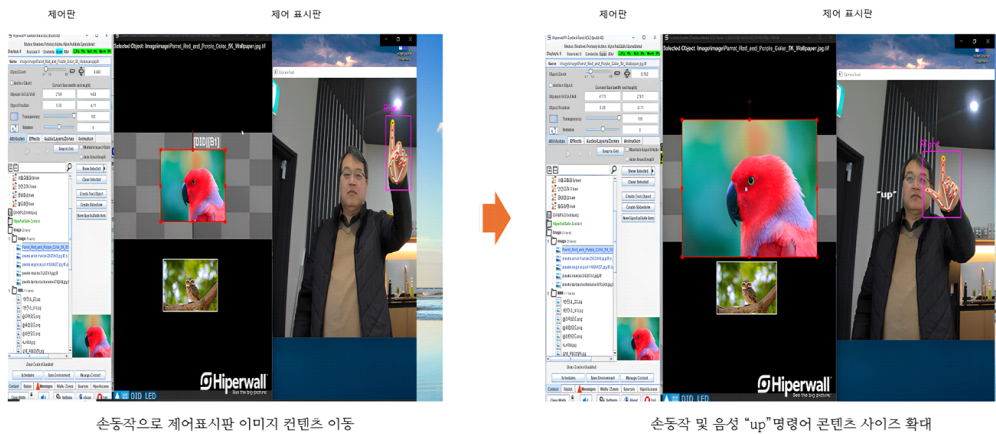
[그림 4-6] 손동작 및 음성명령 제어판의 스크롤바 위로 제어

마지막으로 기능 정의 4와 5에 해당하는 제어표시판에서 콘텐츠를 확대 및 축소하는 기능으로 손동작으로 포인트를 이동하여 콘텐츠를 음성 ‘up’, ‘down’ 명령을 내려 콘텐츠를 확대/축소 기능을 구현하고 테스트하였다. [표 4-7]은 3가지 손동작에 대한 성공률 결과표이다.

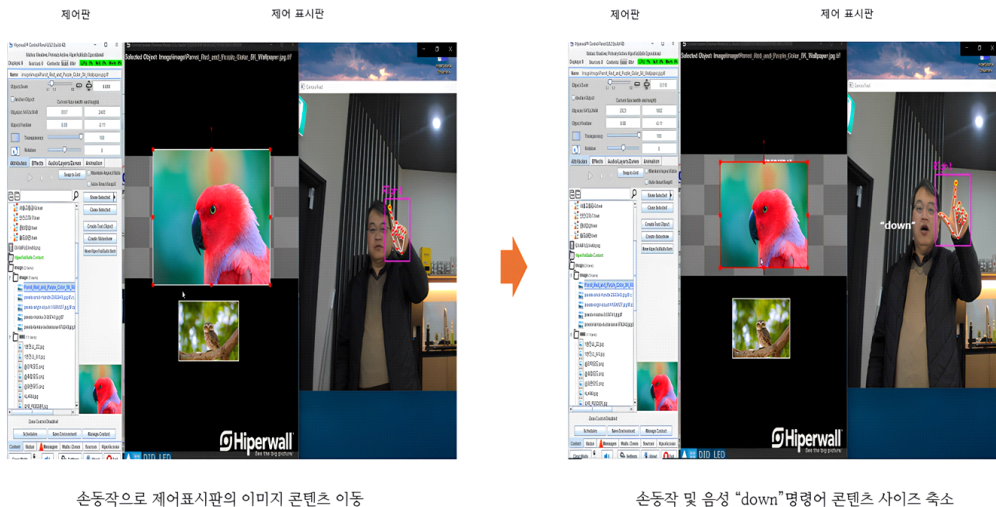
[표 4-7] 3가지 손동작 구현 테스트 결과 5

손동작	테스트 참여자									제어판 기존 제어 방식
	A		B		C		계			
	횏수	성공 횏수	횏수	성공 횏수	횏수	성공 회수	횏수 계	성공 횏수 계	성공률	
 손동작 1	20	20	20	20	20	20	60	60	100%	마우스 휠업 (확대)
 손동작 2	20	18	20	18	20	18	60	54	90%	
 손동작 3	20	20	20	20	20	20	60	60	100%	
 손동작 1	20	19	20	20	20	19	60	58	96.7%	마우스 휠다운 (축소)
 손동작 2	20	20	20	20	20	20	60	60	100%	
 손동작 3	20	20	20	20	20	20	60	60	100%	

[표 4-7]과 같이 마우스 휠업에 대한 기능으로 3가지 손동작을 구현한 결과 손동작 1과 손동작 3이 100% 성공률을 측정되었고, 손동작 2는 90%로 축소하는 기능의 측정치가 낮게 나타났다. 그리고 마우스 휠다운 기능의 손동작으로 손동작 2와 손동작 3이 100%의 성공률로 측정되었고, 손동작 1은 96.7%로 다른 손동작에 비해 낮게 측정되었다. [그림 4-7]과 [그림 4-8]은 제어표시판의 콘텐츠를 손동작 및 음성으로 콘텐츠를 확대/축소하는 기능을 구현하고 테스트한 과정을 캡처한 장면이다.



[그림 4-7] 손동작과 음성 명령 콘텐츠 확대



[그림 4-8] 손동작과 음성 명령 콘텐츠 축소

3장에서 정의한 기능 5가지를 손동작 및 음성으로 콘텐츠 제어하는 기능을 구현하고 테스트를 진행하여 [그림 4-9]는 비디오월에 콘텐츠를 표출하는데 구현하는 테스트 장면이며 자유자재로 표출이 가능한지를 5가지 기능 정의를 기준으로 검증하였다. 다양하고 복잡한 콘텐츠의 기능을 구현하기 위해서는 더 깊이 있는 연구가 필요함을 확인하였다.

[표 4-8] 손동작 및 음성인식 기능정의 테스트 결과

기능 정의 (손동작+음성인식)	테스트 참여자									제어판 기존 제어 방식
	A		B		C		계			
	횟수	성공 횟수	횟수	성공 횟수	횟수	성공 회수	횟수 계	성공 횟수 계	성공률	
기능 정의 1	20	18	20	19	20	19	60	56	93.3%	마우스 클릭/ 더블클릭
기능 정의 2	20	19	20	18	20	19	60	56	93.3%	마우스 좌클릭/ 우클릭
기능 정의 3	20	16	20	17	20	18	60	51	85.0%	마우스 스크롤 업/다운
기능 정의 4	20	17	20	18	20	17	60	52	86.7%	마우스 휠 업 (확대)
기능 정의 5	20	20	20	18	20	17	60	55	91.7%	마우스 휠 다운 (축소)

[표 4-8]과 같이 3장에서 5가지 기능 정의를 통해 각 기능별로 손동작과 음성으로 기능정의 1에서 기능의 5까지 성공률에 대한 기능 테스트를 한 결과로 기능 정의 3이 마우스 스크롤 업다운 기능의 제어 부분에서 85%의 가장낮은 결과가 도출되었고, 기능 정의 1과 기능 정의 2가 93.3%로 나타나 마우스 클릭, 더블클릭, 좌클릭, 우클릭의 기능을 제어하는 부분이 가장 높게 나타났다. 기능 정의 5도 마우스 휠다운(콘텐츠 축소) 기능의 제어로 91.7% 성

공률로 다소 높게 나왔고, 기능 정의 4는 마우스 휠업(콘텐츠 확대) 기능의 제어는 86.7%로 성공률이 다소 낮게 나왔다.



[그림 4-9] 손동작 및 음성으로 비디오월 제어로 콘텐츠 표출

본 연구에서는 기능 정의 5가지를 기준으로 연구 개발하여 검증 및 테스트를 진행한 결과 마우스 다양한 기능을 속도나 정확도 측면에서 손동작 및 음성 인식으로 비디오월 제어하는 부분은 더 많은 연구가 필요하다는 것을 확인하였다.

2) 사용자 평가

가) 평가 개요

비디오월 제어시스템에서 손동작 및 음성으로 마우스 이동, 마우스 좌/우 클릭, 마우스 더블클릭, 마우스 스크롤 업/다운, 마우스 휠업(확대)/다운(축소)의 기능을 구현하고 연구를 사용자의 설문지를 통해 사용자 평가를 통해 본 연구의 사용자 편리성을 검증하고자 하였다. 정보통신 분야에 종사하거나 대학생들 중심으로 10명의 평가단을 구성하여 15문항을 통해 편리성, 만

족도, 활용성 등을 중심으로 평가하였다.

설문은 조사대상자의 응답의 범위를 다양하고 타당성을 높이고 상대적으로 안정된 서열구조를 응답하도록 만든 리커트 척도를 활용하여 조사하였다.

나) 평가단 선정

손동작 및 음성인식으로 비디오 월 제어를 위한 상호 작용연구에서 서울 소재 정보통신 관련 종사자 및 대학생을 10명을 모집하여 진행하였다. 연구 위한 평가단 선정은 협력사 연구원 및 인턴과정의 대학생들에게 연구 목적, 연구장소, 연구 참여에 제한 테스트 및 소요시간을 명시하였다. 신청 후 실험에 연구에 대한 사정 설명 후 서면 동의를 득하고 실시하였다.

다) 평가단 구성

[표 4-9]와 같이 사용자 평가단은 10명 중 남자 4명, 여자 6명으로 구성하여 컴퓨터 7명, ICT 종사자 3명을 대상으로 하였고 연령대는 20대 3명, 30대 3명, 40대 2명, 50대 2명으로 구성하여 설문조사를 실시하였다

[표 4-9] 사용자 평가 구성 정보

문 항	구분	인원	퍼센트(%)
성 별	남	4	40
	여	6	60
전공 및 직업	컴퓨터 공학	7	70
	ICT 분야	3	30
연 령	20대	3	30
	30대	3	30
	40대	2	20
	50대	2	20

라) 마우스와 키보드를 사용과 손동작 및 음성 인식에 대한 느끼는 평가

[표 4-10]과 같이 평소 마우스와 키보드 사용시 불편 정도의 질문에서 불

편함을 없다(50%), 보통이다(40%), 불편하다(10%)로 나타났고, 마우스와 키보드가 불편하다면 어떤 형태의 인터페이스를 선호하는가 평가에서 손동작 및 음성 인식 동시에 인터페이스(60%), 손동작 인식(20%), 음성 인식(20%)로 나타나 손동작 및 음성 인식 동시에 인터페이스가 선호도 가장 높았다.

손동작 및 음성 인식이 동시에 되는 인터페이스가 상용화되려면 어떤 형태를 원하는가에서 운영자 PC설치용 패키지(70%), 플랫폼(20%), 모바일(10%) 순으로 나타났고 운영자 PC에서 패키지로 설치방식이 가장 높게 나왔다. 또한 손동작 및 음성인식이 통합관제센터의 비디오향과 접목시 유용 정도에서 매우유용(50%), 유용(50%)로 나타나 유용성에 높은 평가를 얻을 수 있었다.

[표 4-10] 사용자들이 평소 컴퓨터 상호작용 대한 평가

구 분	많이 불편하다		보통이다		불편함이 없다
마우스 및 키보드 사용 불편한 정도	1(10%)		4(40%)		5(50%)
구 분	손동작		음성인식		손동작 및 음성인식
마우스 및 키보드 불편 대한 인터페이스 선호도	2(20%)		2(20%)		6(60%)
구 분	웹검색		컨텐츠 확대/축소		문서작성
손동작 및 음성 인식 주로 사용 싶은 서비스	2(20%)		2(20%)		6(60%)
구 분	플랫폼		모바일		운영자 PC 설치 패키지용
손동작 및 음성인식 상용화되면 형태	2(20%)		1(10%)		7(70%)
구 분	아주 유용	유용	보통	유용하지 않음	매우 유용하지않음
손동작 및 음성인식 비디오향 유용 정도	5(50%)	5(50%)	0(0%)	0(0%)	0(0%)

그리고 손동작 및 음성 인식이 통합관제센터의 비디오향과 접목시 개선할 점에 대해서는 레이턴시, 인식률, 명령 로딩 줄이기 등에 개선이 필요하다고 평가하였다.

마) 손동작 및 음성 인식 시스템의 기능 구현 만족도 평가

[표 4-11]에서 손동작 및 음성 인식 시스템의 기능 구현 만족도 평가에서 마우스 이동에 손동작 인식은 아주유용(70%), 유용(30%)로 평가하였고 마우스 더블클릭, 마우스 좌/우클릭에 대한 손동작 인식에서 아주 유

[표 4-11] 손동작 및 음성인식 시스템의 기능 구현 만족도 평가

구분	아주 유용	유용	보통	유용하지 않음	매우 유용하지 않음
마우스 이동에 손동작 인식 관련	7(70%)	3(30%)	0(0%)	0(0%)	0(0%)
마우스 더블 클릭에 손동작 인식 관련	5(50%)	5(50%)	0(0%)	0(0%)	0(0%)
마우스 좌클릭에 대한 손동작 인식 관련	5(50%)	5(50%)	0(0%)	0(0%)	0(0%)
마우스 우클릭에 대한 손동작 인식 관련	5(50%)	5(50%)	0(0%)	0(0%)	0(0%)
마우스 스크롤다운에 대한 손동작 인식 관련	3(30%)	6(60%)	1(10%)	0(0%)	0(0%)
마우스 스크롤업에 대한 손동작 인식 관련	3(30%)	6(60%)	1(10%)	0(0%)	0(0%)
마우스 휠업/다운(확대/축소)에 대한 손동작 인식 관련	7(70%)	3(30%)	0(0%)	0(0%)	0(0%)
손동작 인식으로 이미지 및 영상 이동에 관련	5(50%)	5(50%)	0(0%)	0(0%)	0(0%)
마우스 더블 클릭에 음성 인식 관련	2(20%)	8(80%)	0(0%)	0(0%)	0(0%)
마우스 좌클릭에 대한 음성 인식 관련	2(20%)	8(80%)	0(0%)	0(0%)	0(0%)
마우스 우클릭에 대한 음성 인식 관련	2(20%)	8(80%)	0(0%)	0(0%)	0(0%)
마우스 스크롤다운에 대한 음성 인식 관련	2(20%)	2(20%)	5(50%)	1(10%)	0(0%)
마우스 스크롤업에 대한 음성 인식 관련	2(20%)	2(20%)	5(50%)	1(10%)	0(0%)
마우스 휠업/다운 (확대/축소)에 대한 음성 인식 관련	2(20%)	2(20%)	5(50%)	1(10%)	0(0%)

용(50%), 유용(50%)로 평가하고, 마우스 스크롤업/다운에 대해 손동작 인식은 아주유용(30%), 유용(60%), 보통(10%)로 나타나 다소 차이를 보였다. 그리고 마우스 휠업/다운(확대/축소)의 손동작 인식은 아주유용(70%), 유용(30%)로 나타나 유용성을 높게 평가하였다. 음성 인식에 대해서는 마우스 더블 클릭, 마우스 좌/우클릭에 대한 음성 인식에서 아주유용(20%), 유용(80%)로 나타났고, 마우스 스크롤 업/다운, 마우스 휠업/다운(확대/축소)는 아주유용(20%), 유용(20%), 보통(50%), 유용하지않음(10%)로 나타나 다른 기능에 비해 유용성이 떨어지는 것으로 평가하였다.

바) 손동작과 음성 인식의 시스템의 기능 동시 구현 만족도 평가

[표 4-12]은 손동작과 음성 인식을 동시에 구현하였을 때의 기능 만족도 평가로 마우스 이동에 대한 손동작과 음성 인식은 아주유용(70%), 유용(10%), 유용하지 않음(10%), 아주유용하지 않음(10%), 나타나 음성으로 마우스 이동 부분의 한계를 보여주고 있는 것으로 나타났다. 마우스 더

[표 4-12] 손동작과 음성 인식 동시 구현 만족도 평가

구분	아주 유용	유용	보통	유용하지 않음	매우 유용하지 않음
마우스 이동에 손동작 및 음성 인식 관련	7(70%)	1(10%)	0(0%)	1(10%)	1(10%)
마우스 더블 클릭에 손동작 및 음성 인식 관련	5(50%)	4(40%)	1(10%)	0(0%)	0(0%)
마우스 좌클릭에 대한 손동작 및 음성 인식 관련	5(50%)	4(40%)	1(10%)	0(0%)	0(0%)
마우스 우클릭에 대한 손동작 및 음성 인식 관련	5(50%)	4(40%)	1(10%)	0(0%)	0(0%)
마우스 스크롤다운에 대한 손동작 및 음성 인식 관련	1(10%)	4(40%)	5(50%)	0(0%)	0(0%)
마우스 스크롤업에 대한 손동작 및 음성 인식 관련	1(10%)	4(40%)	5(50%)	0(0%)	0(0%)
마우스 휠업/다운(확대/축소)에 대한 손동작 및 음성 인식 관련	4(40%)	4(40%)	2(20%)	0(0%)	0(0%)

블클릭, 마우스 좌/우 클릭에 대한 손동작과 음성 인식은 아주유용(50%, 유용(50%))으로 평가하였고, 마우스 스크롤업/다운은 아주유용(10%), 유용(40%), 보통(50%)로 나타나 보통 이상의 평가를 보였다. 그리고 마우스 휠업/다운(확대/축소)는 아주유용(40%), 유용(40%), 보통(10%)로 대부분 유용하다는 평가를 하였음을 알 수 있었다.

사) 손동작과 음성 인식의 시스템의 기능 동시 구현 시간 만족도 평가

[표 4-13]은 손동작 인식으로 마우스 이동, 더블클릭, 좌/우클릭, 스크롤바 업/다운, 휠업/다운(확대/축소)에 대한 시간은 1.25초에서 1.5초 사이로 만족도가 나타났고, 음성 인식으로 마우스 기능 제어에 대한 부분은 1.25초에서 1.55초 사이도 나타나 손동작 인식과 비슷한 것으로 나타났다. 그리고 손동작과 음성 인식을 동시에 구현하였을 때 1.7초에서 1.8초로 나타나 두가지 동작으로 동시에 하는 것이 기존 마우스로 제어하는 1초이내 보다 늦게 제어가 되는 것으로 나타났다.

[표 4-13] 손동작과 음성 인식 동시 구현 시간 만족도 평가

구분	기존 제어 방식 평균	손동작 및 음성 인식 방식 평균
마우스 이동에 손동작 인식	1초	1.5초
마우스 더블 클릭에 손동작 인식	1초	1.5초
마우스 좌클릭에 대한 손동작 인식	1초	1.5초
마우스 우클릭에 대한 손동작 인식	1초	1.5초
마우스 스크롤다운에 대한 손동작 인식	1초	1.25초
마우스 스크롤업에 대한 손동작 인식	1초	1.25초
마우스 휠업/다운(확대/축소)에 대한 손동작 인식	1초	1.25초
마우스 더블 클릭에 음성 인식	1초	1.25초
마우스 좌클릭에 대한 음성 인식	1초	1.25초
마우스 우클릭에 대한 음성 인식	1초	1.25초
마우스 스크롤다운에 대한 음성 인식	1초	1.35초

마우스 스크롤업에 대한 음성 인식	1초	1,55초
마우스 휠업/다운(확대/축소)에 대한 음성 인식	1초	1.45초
마우스 이동에 손동작 및 음성 인식	1초	1.7초
마우스 더블클릭에 손동작 및 음성 인식	1초	1.7초
마우스 좌클릭에 대한 손동작 및 음성 인식	1초	1.7초
마우스 우클릭에 대한 손동작 및 음성 인식	1초	1.7초
마우스 스크롤다운에 대한 손동작 및 음성 인식	1초	1.8초
마우스 스크롤 업에 대한 손동작 및 음성 인식	1초	1.7초
마우스 휠 업/다운(확대/축소)에 대한 손동작 및 음성 인식	1초	1.8초

3) 연구 성과

본 연구에서 손동작 인식으로 마우스 이동, 더블클릭, 좌/우클릭, 스크롤 업/다운, 휠 업/다운(확대/축소)의 기능을 제어하는 만족도에서 90%이상이 유용하다는 평가가 나왔고, 음성 인식에서는 음성으로 콘텐츠로 이동하는 부분은 제외한 마우스 더블 클릭, 좌클릭, 우클릭의 기능에 대한 제어로 90%이상 유용하다는 나타났고, 마우스 스크롤업/다운, 휠업/다운(확대/축소)의 기능제어는 40%대의 유용하다고 낮은 만족도를 나타냈다. 그리고 손동작과 음성인식을 동시에 비디오월 제어시스템의 UI를 제어할 때 마우스 이동하여 제어하는 기능은 80%의 유용하는 만족도를, 마우스 더블클릭, 좌/우클릭 기능을 손동작과 음성 인식으로 동시에 제어할 때 90%의 유용하다는 높은 만족도를 보였다. 그리고 마우스 스크롤업/다운 기능을 제어할 경우 50%의 유용하다는 결과를 나타내어 다소 낮은 만족도를 보였다. 마우스 휠업/다운(확대/축소) 기능을 제어하는 부분은 80%의 유용하다는 만족도를 나타내었다.

손동작과 음성인식을 동시에 비디오월 제어시스템의 UI를 마우스의 기능을 제어할 경우 기존 마우스보다 0.7초에서 0.8초 늦게 제어되는 것으로 나타나 추가적인 시간을 단축하기 위해 웹캠보다 더 뛰어난 인식 능력을 가진 디바이스로 연구가 진행되어야 한다.

제 5 장 결 론

최첨단 스마트 시티는 도시의 다양한 인프라를 제어하고 모니터링하는 스마트 통합관제센터의 업무의 중요성이 높아지고 있고 통합관제센터의 고도화도 증가하고 있다. 스마트 통합센터를 통해 다양한 도시의 안전 문제를 사전에 인지하고 예방하는 시스템을 인공지능을 활용한 사례가 급증하고 있다. 통합관제센터에 24시간 모니터링을 하는 관제사가 1인당 300개에서 900개의 CCTV 카메라를 모니터링하고 있다. 행정안전부 기준 인당 50대보다 훨씬 초과하는 수치다. 본 연구서에서는 운영 요원의 운영환경을 개선하고자 관제센터의 CCTV를 인간 친화적인 방법으로 운영해 보고자 손동작이나 음성 인식으로 비디오월 제어하고 운영하는 방안을 검토하면서 연구되었다. 인공지능을 이용하여 도시의 다양한 이벤트를 선별하여 관제하는 사례는 응용되고 있지만, 운영자 관점에서 모니터링 컴퓨터와 자연스러운 상호작용을 통해 운영한 사례는 거의 없다. H사에서 통합관제센터인 비디오월 제어시스템을 손동작과 음성 인식 제어가 가능한지를 검토하면서 본 연구를 하게 되었다.

먼저 연구의 기능 정의1에서 손동작과 음성 인식으로 비디오월 제어판에서 제어되는 UI의 폴더를 기존 방식에서 멀티모달 제어가 가능한지에 대한 연구이다. 손동작으로 마우스의 포인트를 이동하여 폴더를 음성명령 ‘Two’를 내려 더블클릭 기능을 수행하여 제어판의 UI를 제어하는 것이다. 기능 구현을 통해 사용자 만족도에서 80% 이상 유용하다는 평가를 얻어 다소 낮은 유용성을 얻었다. 그리고 처리 속도는 1.7초의 측정되어 1초 이하로 높이는 연구가 필요하다.

둘째 기능 정의 2는 손동작과 음성 인식으로 비디오월 제어표시판에서 콘텐츠를 지정하고 마우스 좌클릭 및 우클릭 기능을 활용하여 콘텐츠를 제거하는 기능을 구현하였고 기능 사용성 만족도 조사에서 90%의 유용하다는 평가를 받아 기능의 만족도 나왔다. 99% 이상의 가용성을 올리기 위한 더 많은 연구가 필요함을 확인하였다. 콘텐츠 처리시간도 1.7초로 다소 느리게 반응하였다.

셋째로 기능 정의 3은 손동작과 음성 인식으로 마우스 스크롤바를 위와 아

래로 올리거나 내리는 동작으로 사용자 만족도에 50%의 유용성을 나타내어 가장 낮은 만족도를 나타내었다. 그리고 처리시간도 1.8초로 낮은 처리 시간대를 보여 더 연구의 필요성을 확인하였다.

네 번째와 마지막의 정의 4과 5는 손동작과 음성 인식으로 제어표시판의 콘텐츠의 이미지나 영상을 음성명령을 ‘up’, ‘down’을 내리면 확대되거나 축소되는 기능으로 사용성 평가에서 80%의 유용성을 보여 기능 정의 1과 동일한 수치를 보였다. 처리시간도 1.7초, 1.8초로 나타나 마우스 기능보다는 다소 늦은 동작 반응을 보였다.

본 연구에서는 손동작과 음성 인식의 차원을 넘어 사람처럼 다양한 모달리티로 상호작용하는 기능을 구현하고자 하였고, 구현 대상을 통합관제센터의 모니터링을 위한 비디오월 제어시스템을 대상으로 연구를 진행한 결과 미세한 UI 음향이나 해상도를 조절하기 슬라이드 바의 제어기능을 제외하고 폴더 클릭, 콘텐츠 표출, 콘텐츠 삭제, 콘텐츠 확대 및 축소, 스크롤바 위/아래 제어기능 등은 기능적으로 구현되나 사용자의 만족도에서 기대에 미치지 못한 부분에 연구의 한계가 있었다.

챗 GTP와 같은 거대 언어 모델들이 발전하면서 다양한 형태로 소통하는 세상으로 변화하고 있다. 앞으로 스마트 관제센터의 발전은 다양한 멀티모달 방식을 통해 스마트 통합관제 시스템을 제어하고 모니터링하는 자연 친화적인 시스템의 구현으로 지속적인 연구가 필요하다.

참 고 문 헌

1. 국내문헌

- 강성원, 김철중, 손원. (2009). 손 동작 인식을 통한 사용자에게 편리한 핸드마우스 인터페이스 구현. 『한국방송미디어공학회』, 129-132.
- 김건우, 윤재영. (2021). 인공지능 스피커의 디스플레이 탑재가 사용자 경험에 미치는 영향. Archives of Design Research, 34(3), 173-190..
- 김경수. (2017). DID 복합 콘텐츠 제어를 위한 키넥트 기반 인터페이스. 『한국교육학술정보원』
- 김고운.(2020). Natural User Interface 표준화 동향 연구. 『한국통신학회』, 750-752.
- 김봉환. (2000). 청소년 진로상담 정책의 현황과 과제. 『청소년상담연구』, 8, 146-169.
- 김덕인,서종국.(2022) 광역 스마트 도시 안전망 구축에 관한 연구. 『인천대학교』
- 김도영, 강진아, 이미숙 (2015). 스마트월(Smart-Wall) 원격 다지점 공간의 초실감 가상연결기술. 『전자통신동향분석』, 32(2), 34-4
- 김동욱, 이선경, 정찬양, 이창화, & 백승렬. (2020). 손 위치, 자세, 동작의 통합 심층 학습. 『한국정보처리학회』, 27(2), 1048-1051.
- 김명하, 홍현기.(2013).실시간 동작인식 및 렌더링 기술을 이용한 체감형 콘텐츠 저작도구 개발. 『디지털디자인학연구』,13(3),657-665.
- 김정인. (2012). 손동작 인식을 이용한 마우스제어기법. 『멀티미디어학회』, 15(11),1377-1383.
- 김준우, 맹준희, 주지영, 임광혁.(2010).멀티 터치스크린과 실감형 인터페이스를 적용한 과학 실험 학습 시스템. 『한국콘텐츠학회』,10(8),461-471.
- 김준철. (2021). 확장현실 기반 몰입형 CCTV 스마트 관제 기술. 『서울기술원』, 21(8), 39-41.
- 김종율, 노광현. (2019). 공공 데이터를 활용한 대학생 진로 조언을 위한 그래

- 프레이터베이스기반 지식베이스 구축. 『전자공학회논문지』, 56(10), 41-48.
- 김진형, 김현.(2015).사용자 경험 중심의 내추럴 인터랙션 디자인 방향 연구. 『한국디자인문화학회』, 21(2), 175-186.
- 김희성, 이봉환, & 이규원. (2014). 손동작 인식을 통한 컴퓨터 제어. 『한국정보처리학회』, 21(1), 912-915.
- 김효성, 김병규, 정현기.(2011). 적외선 카메라를 이용한 에어인터페이스 시스템 (AIS) 연구. 『정보처리학회지』, 18(3), 109-116.
- 노정규. (2014). LCD Display 설비 Contents의 Kinect기반 동작제어 기술 구현에 관한 연구. 『전기학회논문지』, 63(4), 565-569.
- 노진우. (2009). 음성 인식을 위한 전처리 방법에 관한 연구. 『국내석사학위논문 고려대학교 대학원』, 서울
- 박강우, 하동환. (2022). LCD 모니터를 활용한 버추얼 스튜디오용 다중 모니터 구현 방법에 관한 연구. 『한국과학예술융합학회』, 40(5), 119-128, 10.17548/ksaf.2022.12.30.119
- 박상운, 이웅주. (2010). 복잡한 영상에 강인한 손동작 인식 방법. 『멀티미디어학회지』, 13(7), 1000-1015.
- 박성도,장복만,장효경,김봉희,최의인.(2010) 멀티모달 인터랙션을 이용한 관제시스템 설계. 『한국통신학회』 : 360-361.
- 박세호, 김태곤, 이지은, & 이정택. (2014). 손 제스처 인식을 위한 손 영역 추적 및 손가락 검출 방법. 『한국방송미디어공학회』, 34-35.
- 박한훈, 최준영, 박종일, & 문광식. (2013). Kinect 기반 손 모양 인식을 위한 손 영역 검출에 관한 연구. 『방송공학회논문지』, 18(3), 393-400.
- 성민정. (2020). 실감형 인터랙션 미디어 환경의 경험적 가치향상을 위한 프레즌스 요소에 관한 연구. 『한양대학교 대학원』, 서울
- 송락빈, 홍윤아, 곽노윤. (2023). MediaPipe Hands 모델 기반의 손 제스처 인식을 이용한 사용자 인터페이스. 『멀티미디어학회논문지』, 26(2), 103-115.
- 양선진, 박재표, & 양승민. (2019). 스마트 시티 환경에서 대규모 영상 모니터링

- 을 위한 IP 비디오 월 시스템의 설계 및 구현. 『한국산학기술학회 논문지』, 20(9), 7-13.
- 여성수, 김재원, 이재용. (2021). 도시 환경에서의 디지털트윈 물리보안 시스템 논의. 『한국경찰학회』, 23(6), 153-177.
- 유건식. (2020). 코로나19 시대와 방송 제작 방식의 변화. 『방송과미디어』, 25(4), 51-60.
- 유인호, 신동윤. (2012). Easyoloy 학습조차 필요없는 편안함 추구해야. 『동아비즈니스리뷰』.
- 유현수.(2021). 새로운 미래의 인터페이스 기술, NICE평가정보(주).
- 이대검, 김원, 윤여경.(2020). NUI를 활용한 가상현실 박물관 전시 디자인 연구. 『한국상품문화 디자인학회』, 62,99-109.
- 이광형, 신동규 and 신동일. (2014). 직관적인 핸드 모션에 기반한 NUI/NUX 프레임워크. 『인터넷정보학회논문지』, 15(3), 11-19.
- 이동우, 손용기, 김지은, 선우존, 조일연, 한동원. (2007). 웨어러블 컴퓨터를 위한 음성 제스처 기반 멀티모달 융합 사용자 인터페이스. 『한국차세대컴퓨팅학회』, 3(2), 17-25.
- 이래경,김성신.(2001). 손동작 인식을 통한 Human-Computer 구현. 『한국지능시스템학회』 11(1),28-32.
- 이민규, 전재봉.(2012).[우수논문] 키넥트를 이용한 개인용 컴퓨터 제어. 『한국정보과학회 학술발표논문집』,39(1A),343-345.
- 이윤근. (2012). 음성인터페이스 기술 개요 및 스마트폰 환경에서의 서비스 동향. 『한국통신학회지(정보와통신)』, 29(4), 3-9.
- 이정진, 김종호,김태영. (2010). 증강현실 응용을 위한 손끝점 추출과 손동작 인식 기법. 『멀티미디어학회지』, 13(2), 316-323.
- 이창주, 이준호, (2007) “깊이 예지 기반의 Curvature Scale Space Map을 이용한 손 제스처 인식”, 『한국정보처리학회』 제14권 제1호
- 이하은, 한정협. (2021). 메타버스 플랫폼의 체험형태에 따른 유형 분류 및 특성 연구:실감, 초실감 메타버스 대표사례를 중심으로. 『한국공간디자인학회』, 16(8), 427.

- 임미정. (2006). 멀티모달 인터페이스 개발을 위한 휴먼-컴퓨터 인터랙션 설계. 『아주대학교』
- 임미정, & 박범. (2006). 멀티모달 인터랙션을 위한 사용자 병렬 모달리티 입력 방식 및 입력 동기화 방법 설계. 『대한인간공학회지』, 25(2), 135-146.
- 장복만, 김정석, 장효경, 박성도, 최의인.(2009). 멀티모달 인터페이스를 이용한 관제시스템. 『한국정보기술학회』, (6),934-938.
- 정석용, 박형수, 강환수, 조진형, Larry S Kim.(2023).메타버스 기반 실감 교육용 플랫폼 개발. 『한국컴퓨터정보학회』, 28(2),77-87. 재인용
- 전창의, 이효정, 김기범. (2020). 음성 AI시장의 동향과 비즈니스 기회. 『삼정 KPMG 경제연구원 ISSUE MONITOR』, 제126호.
- 장일식, 박종철. (2018). CCTV 통합관제센터의 운영성과 및 개선 방안-구미시를 중심으로 -. 『경찰학논총』, 13(1), 209-246.
- 정유경. (2011). 멀티 디스플레이 사례분석과 관리 인터페이스에 관한 연구. 『디자 인지식저널』, 18, 180-189.
- 정지인, 김혜림, 서세현, 계이기, 최황규. (2009). 손동작 인식을 이용한 프로그램 제어 인터페이스 설계 및 구현. 『한국정보처리학회지』, 16(2), 383-384
- 정충식. (2015). CCTV 통합관제센터의 운영 개선 방안에 관한 연구: 부산시 통합관제센터를 중심으로. 『한국지역정보화학회지』, 18(3), 123-154.
- 조영욱, 노광현. (2023). IP 비디오 Wall 기반 실시간 양방향 모바일 방송시스템에 관한 연구. 『아시아태평양융합연구교류논문지』, 9(2), 1-12.
- 조운상, 이지윤, 이영재, 강현규, 김성범.(2019).Multimodal Deep Learning for Explainable Product Design Data. 『대한산업공학회 추계학술대회 논문집』, 2019(11) 1184-1216.
- 최원석, 이동우, 김문식, 나종화.(2007).음성인식 및 영상처리 기반 멀티모달 입력 장치의 설계. 『제어로봇시스템학회』, 13(8),743-748.
- 최광국, 곽상훈, 김유진, 김철, & 최승호. (2000). 멀티모달을 이용한 응용프로그램 제어에 관한 연구. 『한국음향학회』, 205-208.

하유솔, 고은지, 김명준.(2011). 손동작 인식 기반의 마우스 입력 장치. 『한국정보처리학회』, 38(10),524-535.

홍기형. (2006). XML 기반 멀티모달 상호작용 표준 언어. Korea Multimedia Society, 10(1), 71-82.

한충희, 한창희. (2021). 안전한 스마트월드를 위한 사이버 테러위협 제거 방안 연구. 『융합보안논문지』, 21(1), 107-113.

그레이트포춘2유. <https://greatfortune2u.tistory.com/28>

네이브 블로그. <https://m.blog.naver.com/vinylx/220367088474>

네이버. <https://blog.naver.com/mosfnet/221988404492>

네이버. <https://blog.naver.com/msnayana/80199112243>

네이버. <https://blog.naver.com/msnayana/80102408151>

동아. https://dbr.donga.com/article/view/1101/article_no/4990

더블유쓰리. <http://www.w3.org/TR/2003/NOTE-mmi-reqs-20030108/>

디벨로퍼즈. <https://developers.google.com/mediapipe/solutions>

디벨리퍼어라이스. <https://developeralice.tistory.com/10>

디데일리. <https://ddaily.co.kr/m/page/view/2023122915351101186>

러닝믹스. <https://blog.learningbix.com/creating-ai-virtual-mouse/>

밴드리코. <https://vandrico.com/wearables/device/logbar-ring.html>

사이드웨어. <https://www.slideshare.net/JonghoonSeo/20-24091955>

아이셈. <https://ko.isemc.com/news-isemc/what-is-video-wall>

아이셈. <https://ko.isemc.com/news-isemc/what-is-a-video-wall-controller>

애플. <https://www.apple-economy.com/news/articleView.html?idxno=70290>

에트리. <https://www.etri.re.kr/webzine/20190215/sub01.html>

에이아이타임즈. <https://www.aitimes.kr/news/articleView.html?idxno=25562>

엠디에스테크. <https://www.mdstech.co.kr/reference/4934>

에이아이타임즈. <https://www.aitimes.kr/news/articleView.html?idxno=25795>

에이에프엔씨. <https://www.epnc.co.kr/news/articleView.html?idxno=13427>

에이에프엔씨. <https://www.epnc.co.kr/news/articleView.html?idxno=13427>

위키피디아. https://ko.wikipedia.org/wiki/음성_인식

위키피디아시큐리티. <https://wiki.wikisecurity.net/identification>: 음성인식
이티뉴스. <https://www.etnews.com/20220919000157>
제스처테크. <https://www.gesturetek.com/>
제스처테크헬쓰. <https://gesturetekhealth.com/products/gestrack-3d>
콜센터. <https://www.callcenter.or.kr>
컴퓨팅. <http://computing.or.kr/14792/speech-recognition> 음성-인식/
카드. <https://www.kidd.co.kr/news/216503>
한인포스트. <https://haninpost.com/archives/89763>
허깅페이스. <https://huggingface.co/MIT/ast-finetuned-speech-commands-v2>
cctv 뉴스. <https://www.cctvnews.co.kr/news/articleView.html?idxno=215145>

2. 국외문헌

- Benoit, C., Martin, J. C., Pelachaud, C., Schomaker, L., & Suhm, B. (2000). Audio-visual and multimodal speech systems. Handbook of Standards and Resources for Spoken Language Systems-Supplement, 500, 1-95.
- Bhanderi, Dinesh J. Mishra, Daxa G. Parikh, Shweta M. Sharma, Deepak B. (2017). Computer Use and Carpal Tunnel Syndrome: A Case-control Study. Indian Journal of Occupational & Environmental Medicine. Vol. 21 Issue 3, p109-114. 6p. 6 Charts.
- Cao, C., Sun, Y., Li, R., & Chen, L. (2011). Hand posture recognition via joint feature sparse representation. Optical Engineering, 50(12), 127210-127210.
- Choi, J., Park, H., & Park, J. I. (2011). Hand shape recognition using distance transform and shape decomposition. International Conference on Image Processing (pp. 3605-3608). IEEE.
- Choi, C. S., & Oh, K. S. (2010). Gesture Recognition Algorithm from the Webcam Images. In *Proceedings of the KAIS Fall Conference* (pp. 88-91). The Korea Academia-Industrial cooperation Society.
- Gesture Internet Document (2007) From “<http://www.en.Wikipedia.org/wiki/Gesture>”
- Gong, Y., Chung, Y. A., & Glass, J. (2021). Ast: Audio spectrogram transformer. arXiv preprint arXiv:2104.01778.
- Lahiani, H., & Neji, M. (2018). Hand gesture recognition method based on HOG-LBP features for mobile devices. Procedia Computer Science, 126, 254-263.
- Modler, P., & Myatt, T. (2008). Recognition of separate hand gestures by time-delay neural networks based on multi-state spectral image patterns from cyclic hand movements. In 2008 IEEE International

- Conference on Systems, Man and Cybernetics (pp. 1539–1544).
- J.Ohya and Y.Kitamura, etc.(1993). Real-Time Reproduction of 3D Human Images in Virtual Space Teleconferencing, in Proc. 『Virtual Reality Annual Int. Symp』 . 408–414,
- Jin, S., Li, Y., Lu, G. M., Luo, J. X., Chen, W. D., & Zheng, X. X. (2011). SOM-based hand gesture recognition for virtual interactions. In 2011 IEEE International Symposium on VR Innovation (pp. 317–322).
- Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., ... & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal processing magazine*, 29(6), 82–97.
- Huang, X., Acero, A., Hon, H. W., & Reddy, R. (2001). Spoken language processing: A guide to theory, algorithm, and system development. Prentice hall PTR.
- I. Oikonomidis, N. Kyriazis, and A.A.(2011) Argyros,"Efficient Model-based 3D Tracking of Hand Articulations using Kinect," Proc. of British Machine Vision Conf., pp. 101.1–101.11.
- Kavana, K. M., & Suma, N. R. (2022). Recognition of Hand Gestures Using Mediapipe Hands. *International Research Journal of Modernization in Engineering Technology and Science*, 4(06).
- Mehrish, A., Majumder, N., Bharadwaj, R., Mihalcea, R., & Poria, S. (2023). A review of deep learning techniques for speech processing. *Information Fusion*, 101869.
- Norris, S. (2006). Multiparty interaction: a multimodal perspective on relevance. *Discourse Studies*, 8(3), 401–421..
- Pinola, M. (2017). History of voice recognition: from Audrey to Siri. itbusiness. ca.

- Rabiner, L. R., & Juang, B. H. (2006). Speech recognition: Statistical methods. *Encyclopedia of language & linguistics*, 1–18.
- Sang-Hun Lee, Dae-jin Kim, & Hong-Sub Choi (2012). Implementation of DID interface using gesture recognition. *Journal of Digital Contents Society*, 13(3), 343–352
- Shriram, S., Nagaraj, B., Jaya, J., Shankar, S., & Ajay, P. (2021). Deep learning-based real-time AI virtual mouse system using computer vision to avoid COVID-19 spread. *Journal of healthcare engineering*, 2021
- Trindade, P., Lobo, J., & Barreto, J. P. (2012). Hand gesture recognition using color and depth images enhanced with hand angular pose data. In *2012 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)* (pp. 71–76). IEEE.
- Vinciarelli, Alessandro(2009). Social signal processing: Survey of an emerging domain. In *Image and Vision Computing*, 27(12), 1743–1759
- Wu Y, Huang TS (1999) Vision-based gesture recognition: a review. *International Gesture Workshop*, France.
- Xu, H., Hou, X., Su, R., & Ni, Q. (2009). Real-time hand gesture recognition system based on associative processors. In *2009 2nd IEEE International Conference on Computer Science and Information Technology* (pp. 14–18).
- Yun, L., & Peng, Z. (2009). An automatic hand gesture recognition system based on Viola-Jones method and SVMs. In *2009 Second international workshop on computer science and engineering* (Vol. 2, pp. 72–76).
- Zahra, R., Shehzadi, A., Sharif, M. I., Karim, A., Azam, S., De Boer, F., ... & Mehmood, M. (2023). Camera-based interactive wall display using hand gesture recognition. *Intelligent Systems with Applications*, 19, 200262.

손동작 및 음성 인식의 사용 편의성/만족도 설문지

안녕하십니까?

본 조사는 일반 컴퓨터 공학 전공 대학생 및 ICT분야에 종사하는 일반인을 대상으로 손동작 및 음성 인식 인터페이스에 대한 사용 편의성을 조사하기 위한 설문입니다.

주어진, 이용 안내서를 따라서 개발된 손동작 및 음성인식 시스템을 이용해 보신 후 사용 편의성, 만족도에 대해 솔직하게 표기를 해 주시기 바랍니다.

응답해 주신 모든 분들의 개인정보는 통계법 제 13조에 의해 비밀이 보호됩니다.

■ 응답자 정보 및 진로 관련 내용

문 1. 귀하의 성별은?

- 1) 남 2) 여

문 2. 귀하의 전공 및 직업은?

- 1) IT융합공학 2) 컴퓨터공학 3) ICT 종사 4) 기타 ()

문 3. 귀하의 연령대는?

- 1) 20대 2) 30대 3) 40대 4) 50대이상

문 4. 현재 귀하의 평소 컴퓨터를 사용할 때 마우스나 키보드에 대한 불편함을 얼마나 느끼나요?

- 1) 많은 불편함을 느낀다 2) 보통이다 3) 불편함을 느끼지 못한다.

문 5. 귀하는 마우스나 키보드 사용시 불편함을 느낀다면 어떤 형태의 인터페이스 선호하시나요?

- 1) 손동작 인식 2) 음성 인식 3) 손동작과 음성 인식 기반 멀티모달 인식

문 6. 귀하는 마우스나 키보드와 손동작이나 음성을 인식하는 인터페이스 사용시 주로 사용하고 싶은 서비스는?

- 1) 웹 검색 2) 이미지/영상 자유자재로 확대, 축소 및 이동 3) 문서 작성/편집

문 7. 향후 손동작 및 음성외 추가로 인터페이스 되는 부분이 있다면? (복수 가능)
()

문 8. 손동작 및 음성 인식이 동시에 되는 인터페이스가 상용화되면 어떤 형태를 원
하는가?

1) 플랫폼 서비스 2) 모바일 앱 3) 일반 웹 4) 운영자 PC 설치용 패키지

문 9. 손동작 및 음성 인식이 통합관제센터의 비디오월과 접목시 유용하게 사용될
것으로 보는가?

1) 아주 유용 2) 유용 3) 보통이다 4) 유용하지 않다 5) 매우 유용하지 않다

문 10. 손동작 및 음성 인식이 통합관제센터의 비디오월과 접목시 개선해야 될 사
항이 있다면?
()

■ 손동작 및 음성 인식 시스템의 기능 구현/만족도 평가

문 11. 테스트하신 손동작 인식 시스템의 기능에 각 질문의 답변에 대한 만족도를 평
가해 주시기 바랍니다. (5점 척도)

구 분	매우 유용 하다 5	유용하다 4	보통이다 3	유용하지 않다 2	전혀 유 용하지 않다 1
1. 마우스 이동에 대한 손동작 인식	5	4	3	2	1
2. 마우스 더블클릭에 대한 손동작 인식	5	4	3	2	1
3. 마우스 좌클릭에 대한 손동작 인식	5	4	3	2	1
4. 마우스 우클릭에 대한 손동작 인식	5	4	3	2	1
5. 마우스 스크롤 다운에 대한 손동작 인식	5	4	3	2	1
6. 마우스 스크롤 업에 대한 손동작 인식	5	4	3	2	1
7. 마우스 휠 업(확대)에 대한 손동작 인식	5	4	3	2	1
8. 손동작 인식으로 이미지 및 영상 이동	5	4	3	2	1

문12. 테스트하신 음성 인식 시스템의 기능에 각 질문의 답변에 대한 만족도를 평가해 주시기 바랍니다. (5점 척도)

구 분	매우 유용 하다 5	유용하다 4	보통이다 3	유용하지 않다 2	전혀 유용 하지 않다 1
1. 마우스 이동에 대한 음성 인식	5	4	3	2	1
2. 마우스 더블클릭에 대한 음성 인식	5	4	3	2	1
3. 마우스 좌클릭에 대한 음성 인식	5	4	3	2	1
4. 마우스 우클릭에 대한 음성 인식	5	4	3	2	1
5. 마우스 스크롤 다운에 대한 음성 인식	5	4	3	2	1
6. 마우스 스크롤 업에 대한 음성 인식	5	4	3	2	1
7. 마우스 휠 업(확대/축소)에 대한 음성 인식	5	4	3	2	1

문13. 테스트하신 손동작 및 음성 인식 시스템의 기능에 각 질문의 답변에 대한 만족도를 평가해 주시기 바랍니다. (5점 척도)

구 분	매우 유용 하다 5	유용하다 4	보통이다 3	유용하지 않다 2	전혀 유용 하지 않다 1
1. 마우스 이동에 손동작 및 음성 인식	5	4	3	2	1
2. 마우스 더블클릭에 손동작 및 음성 인식	5	4	3	2	1
3. 마우스 좌클릭에 손동작 및 음성 인식	5	4	3	2	1
4. 마우스 우클릭에 손동작 및 음성 인식	5	4	3	2	1
5. 마우스 스크롤 다운에 손동작/음성 인식	5	4	3	2	1
6. 마우스 스크롤 업에 손동작 및 음성 인식	5	4	3	2	1
7. 마우스 휠 업/다운(확대/축소)에 손동작 및 음성 인식	5	4	3	2	1

문14. 손동작 및 음성 인식 시스템 구현시 불편한 사항/추가 필요한 사항/문제되는 사항을 기입하여 주시기 바랍니다.

()

문15. 아래의 정보를 기존 방식(마우스나 키보드)와 손동작 및 음성 인식 시스템을 활용하여 구현 시간을 입력해 주시기 바랍니다.

(1) 손동작 인식

구 분	기존 (키보드 및 마우스)	손동작 인식
1. 마우스 이동 손동작 인식		
2. 마우스 더블클릭 손동작 인식		
3. 마우스 좌클릭 손동작 인식		
4. 마우스 우클릭 손동작 인식		
5. 마우스 스크롤 다운 손동작 인식		
6. 마우스 스크롤 업 손동작 인식		
7. 마우스 휠 업/다운(확대/축소) 손동작 인식		
8. 손동작 인식으로 이미지 및 영상 이동		
전체 만족도(백분률)		

(2) 음성 인식

구 분	기존 (키보드 및 마우스)	손동작 인식
1. 마우스 이동 음성 인식		
2. 마우스 더블클릭 음성 인식		
3. 마우스 좌클릭 음성 인식		
4. 마우스 우클릭 음성 인식		
5. 마우스 스크롤 다운 음성 인식		
6. 마우스 스크롤 업 음성 인식		
7. 마우스 휠 업/다운(확대/축소) 음성 인식		
전체 만족도(백분률)		

(3) 손동작 및 음성 인식

구 분	기존 (키보드 및 마우스)	손동작 인식
1. 마우스 이동 손동작 및 음성 인식		
2. 마우스 더블클릭 손동작 및 음성 인식		
3. 마우스 좌클릭 손동작 및 음성 인식		
4. 마우스 우클릭 손동작 및 음성 인식		
5. 마우스 스크롤 다운 손동작 및 음성 인식		
6. 마우스 스크롤 업 손동작 및 음성 인식		
7. 마우스 휠 업/다운(확대/축소) 손동작 및 음성 인식		
전체 만족도(백분률)		

ABSTRACT

A Study on Multimodal Interaction Using Hand Gestures and Speech for Video Wall Control System

Cho, Young-Wook

Major in Smart Convergence Product

Dept. of Smart Convergence Consulting

The Graduate School

Hansung University

With the development of the display industry, the part that interacts with computers through various interfaces is increasing. It is possible to control the computer with hand gestures or speech without direct control of a mouse or keyboard, or to listen to music or guide directions with hand gestures and voice commands in a car. Voice and hand motion recognition technology are being applied to various applications in entertainment, medical care, robotics, and smart cities. The importance of smart cities in which autonomous vehicles, autonomous security robots, and autonomous delivery robots walk through cities, and drone taxis fly in the sky is increasing. The function of the smart integrated control center that monitors and controls various problems in the city is also being advanced. Monitoring of safety not only in cities but also in railroads, power plants, manufacturing, and construction has been strengthened, and the enforcement of the Serious Accident Act made the management and supervision of business owners more severe.

This study aims to efficiently monitor the control center operator by developing a function to interact in a nature-friendly way with the video wall control system, a system that monitors and controls CCTV images of smart integrated control systems. Web cameras (RGB), 3D cameras, Leap Motion, and LiDAR interface equipment are used to recognize hand gestures, and in this study, a widely distributed web camera (RGB) was used, and a microphone was used as an interface device for speech recognition. The hand gestures and speech detected by the input device are designed to be controlled by sequentially synchronizing and interacting with the hand gestures and speech signals after actual detection verification through preprocessing.

The feature-specific technical elements for voice and hand-shaped recognition of the video wall control system using a multimodal interface are constructed, and hand-shaped recognition is implemented using Google's machine learning-based high-performance hand gesture recognition solution, MediaPipe, through webcam. 21 3D landmarks are deduced for each hand by frame to recognize two or more hands at the same time and interface. After recognizing the hand shape, the UI of the control panel of the video wall was controlled through a voice command to display content such as images and web browsers on the video wall, and functions of deleting, enlarging, and reducing content were developed to implement a nature-friendly video wall control system.

In order to apply the most appropriate hand gesture, three hand gestures were tested, and each hand gesture was selected by verifying the accuracy and convenience of recognition. Speech recognition is designed by combining the model of 'MIT/ast-defined-speech-commands-v2' so that when a voice command is given, the voice is converted into text and controlled in conjunction with the hand gesture at the same time.

It was applied to the designed hand motion and voice recognition-based

video wall control system, and used and tested for computer engineering students and ICT workers for verification of the system, compared to the existing mouse control method, and verified and evaluated for satisfaction and usefulness through user evaluation. Among the functions of controlling the video wall control system with hand motion and voice recognition, left-click and right-click functions of the mouse showed the highest usefulness of 90%, and mouse movement, double-click, and wheel up/down(zooming/down) were slightly lower at 80%. And the control of the mouse scroll up/down function confirmed the difficulty of controlling the scroll bar at the lowest 50%, and when the video wall control system controlled the UI of the control panel with hand motion and voice recognition, the motion was recognized 0.3 to 0.55 seconds faster than when controlled sequentially with hand motion and voice recognition.

The results of this study confirmed the possibility and usefulness of controlling the video wall control system by hand motion or voice based on multimodal. In addition, in order to recognize hand motion, it has become recognized that it is necessary to research and develop a platform that is recognized by various devices such as LiDAR and LeapMotion from a general-purpose device such as a webcam.

In addition, after issuing the command with hand motion and voice recognition, it was confirmed that the response time of the content must be similar to or faster than the function of the mouse to control the content. In particular, it is judged that research that controls the computer GUI environment in a more precise way should be conducted to control the slide bar that controls sound and resolution among the UI of the control panel.

【Key word】 multimodal, voice recognition, hand gestures recognition, smart control center, video wall control