

감정 인식을 활용한 스마트홈 자동화 시스템

3반 3팀

김기수 김두성 이형욱

1. 소개

현대 사회에서는 인공지능과 데이터 과학이 발전하면서 사용자에게 간편한 맞춤형 서비스가 이미 일상 속에 들어와 있습니다. 이에 따라 사용자가 직접 제어하지 않고도 사용자의 상태를 자동으로 파악해 편리한 스마트홈 환경을 제공하려고 합니다.

전자기기에 익숙한 20~30대 사용자가 원하는 대로 커스텀하여 사용할 수 있고 전자기기에 익숙하지 않은 사용자도 쉽게 사용할 수 있도록 제어가 필요없는 기본적인 자동화를 제공합니다.

라즈베리파이3에 카메라를 연결하여 홈 CCTV를 구현하고, 사용자의 감정을 인공지능을 통해 파악하고 연결된 집의 가전들을 자동으로 제어합니다.

2. 요구사항

<기능 요구사항>

1. 웹캠을 통해 사용자의 얼굴을 실시간으로 감지합니다.
2. 감지된 얼굴을 시각화하여 볼 수 있습니다.
3. 감정 인식 모델을 통해 사용자의 현재 감정 상태를 파악하여 모니터에 보여줍니다.
(화, 짜증, 두려움, 행복, 슬픔, 놀람, 중립 총 7개)
4. 인식된 감정에 따라 조명의 색상 및 밝기를 조절합니다.
5. 인식된 감정에 따라 어울리는 음악을 재생합니다.
6. AI 스피커가 사용자의 감정에 따른 대화를 시도합니다.

<비 기능 요구사항>

감정을 활용할 수 있는 모듈 추가가 용이해야 합니다.
사용자의 개인 정보 활용 시 명시해야 합니다.

* 제약사항:

- 웹캠이 연결되어 있어야 합니다.
- 인터넷 연결이 필요합니다.
- 라즈베리파이3에서 사용 가능합니다.

* 품질 속성:

- 감정 인식 정확도는 테스트 데이터 기준 60% 이상 이어야 합니다.
- 단위 프레임당 얼굴 감지 및 감정 추론이 0.3초 이내에 완료되어야 합니다.

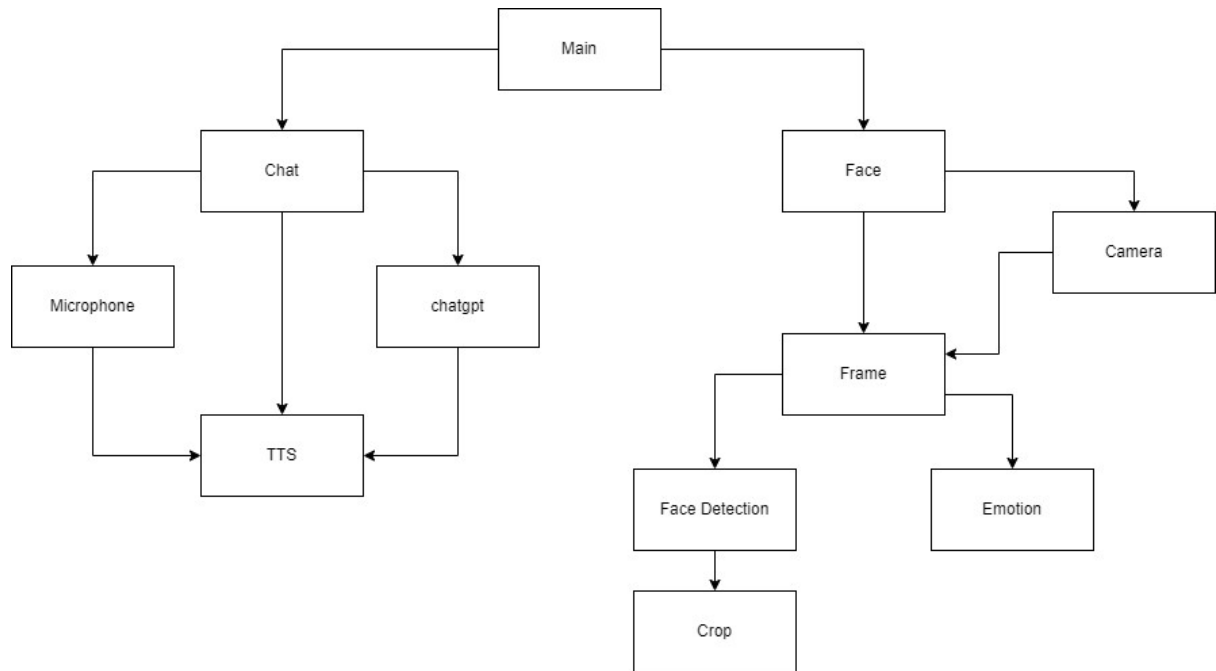
3. 설계

<prototyping 모델>

빠르게 프로토타입을 만들고, 이를 바탕으로 지속적인 개선을 통해 최종 제품을 완성합니다.
독립적인 모듈이 작동하는 프로토타입을 만든 후, 서로의 모듈을 결합합니다.

1. 노트북으로 웹캠을 통해 사용자의 얼굴을 인식하고 감정을 추출
 - 감정 인식률을 높일 수 있도록 사용자의 얼굴 이미지 전처리 추가
 - 라즈베리파이에서도 작동하는 지 확인 필요
2. 임시로 추출된 감정에 따라 AI 대화
 - 대화 말투 변경
 - 대화를 끝마치는 방식 논의
3. 라즈베리파이3에서 웹캠을 통해 얼굴을 인식하고 감정 추출 후 AI 대화
 - 속도 개선 및 AI 모델 수정
4. AI대화에 조명 제어, 음악 추천 기능 추가
 - 활용 효과를 직관적으로 알 수 있게 데모 필요

<모듈 뷰>



1. Main 모듈

전체 시스템의 동작을 제어합니다. 다른 모듈들이 원활하게 연동되도록 조정하고, 시스템의 동작 순서를 관리합니다.

2. Face 모듈

사용자의 얼굴을 인식하는 역할을 합니다. 카메라로부터 입력된 이미지에서 얼굴을 찾아내는 기능을 담당합니다.

3. Camera 모듈

카메라를 제어하여 사용자의 얼굴 이미지를 캡처하는 역할을 합니다. 캡처된 이미지는 Frame 모듈로 전달됩니다.

4. Frame 모듈

카메라에서 캡처된 이미지를 처리하여, 얼굴 인식에 적합한 형태로 변환하는 역할을 합니다.

5. Face Detection 모듈

Frame 모듈에서 전달받은 이미지 내에서 얼굴을 정확하게 탐지하는 역할을 합니다. 탐지된 얼굴 정보는 Crop 모듈로 전달됩니다.

6. Crop 모듈

Face Detection 모듈에서 탐지된 얼굴을 이미지에서 잘라내는 역할을 합니다. 잘라낸 얼굴 이미지는 Emotion 모듈로 전달됩니다.

7. Emotion 모듈

Crop 모듈에서 전달받은 얼굴 이미지를 분석하여 사용자의 감정 상태를 판단하는 역할을 합니다.

8. Chat 모듈

사용자와 시스템 간의 대화를 제어하는 역할을 합니다. Emotion 모듈에서 판단한 감정 상태에 따라 적절한 대화 스크립트를 선택하고, 이를 TTS 모듈로 전달합니다.

9. Microphone 모듈

사용자의 음성 입력을 받아서 텍스트로 변환하는 역할을 합니다. 변환된 텍스트는 Chat 모듈로 전달됩니다.

10. ChatGPT 모듈

사용자와의 대화를 자연스럽게 이끌어가는 역할을 합니다. Microphone 모듈에서 전달받은 사용자의 음성 입력을 바탕으로 적절한 응답을 생성하고, 이를 TTS 모듈로 전달합니다.

11. TTS(Text-To-Speech) 모듈

텍스트를 음성으로 변환하는 역할을 합니다. Chat 모듈이나 ChatGPT 모듈에서 전달받은 텍스트를 음성으로 변환하여 사용자에게 전달합니다.

4. 검증 테스트

1. 카메라를 통해 영상이 모니터에 실시간으로 잘 나오는 지 확인합니다.
2. 모니터 영상에서 얼굴 부분을 인식하여 표시합니다.
3. 7가지 감정이 드러나도록 표정을 짓고 해당 표정을 인식합니다.
4. 각각의 감정에 맞는 제어가 되는 지 확인합니다.
 - 4-1. 감정에 맞는 음악 재생 (부록1 참조)
 - 4-2. 감정에 맞는 조명 제어 (부록2 참조)
 - 4-3. 감정에 맞는 AI 대화 (AI가 사용자의 감정을 말하고 그에 맞는 대화를 합니다)

5. 개발자 테스트

<단위 테스트>

각 모듈의 기능을 개별적으로 테스트하여 해당 모듈이 정확하게 작동하는지 확인합니다.

1. Main 모듈 검증

Main 모듈은 다른 모듈들의 동작을 조정하므로, 다른 모듈들이 정상적으로 작동하면서 Main 모듈의 명령에 따라 동작하는지를 확인합니다.

2. Face 모듈 검증

Face 모듈은 다양한 얼굴 이미지를 입력으로 받아, 얼굴 인식이 정확하게 이루어지는지 검증합니다. 이를 위해 다양한 각도와 표정, 조명 조건에서의 얼굴 이미지를 사용합니다.

3. Camera 모듈 검증

Camera 모듈은 카메라의 작동 상태와 이미지 캡처 기능을 검증합니다. 카메라가 적절한 시간에 활성화되고 이미지를 캡처하는지 확인하는 방식으로 검증합니다.

4. Frame 모듈 검증

Frame 모듈은 입력 이미지를 처리하는 기능을 검증합니다. 원본 이미지와 처리된 이미지를 비교하여, 이미지 처리가 정확하게 이루어지는지 확인합니다.

5. Face Detection 모듈 검증

Face Detection 모듈은 얼굴 탐지 기능을 검증합니다. 다양한 이미지에서 얼굴이 정확하게 탐지되는지 확인하며, 얼굴이 없는 이미지에서는 얼굴이 탐지되지 않는지 확인합니다.

6. Crop 모듈 검증

Crop 모듈은 얼굴 이미지의 잘라내기 기능을 검증합니다. 잘라낸 이미지가 얼굴 부분만을 정확하게 포함하고 있는지 확인합니다.

7. Emotion 모듈 검증

Emotion 모듈은 감정 분석 기능을 검증합니다. 다양한 표정을 가진 얼굴 이미지를 입력으로 하여 감정 분석 결과가 정확한지 확인합니다.

8. Chat 모듈 검증

Chat 모듈은 대화 제어 기능을 검증합니다. 감정 상태에 따라 적절한 대화 스크립트가 선택되는지 확인합니다.

9. Microphone 모듈 검증

Microphone 모듈은 음성 인식 기능을 검증합니다. 다양한 음성 입력을 제공하여, 인식된 텍스트가 정확한지 확인합니다.

10. ChatGPT 모듈 검증

ChatGPT 모듈은 대화 생성 기능을 검증합니다. 다양한 입력 텍스트를 제공하여, 생성된 응답이 적절한지 확인합니다.

11. TTS(Text-To-Speech) 모듈 검증

TTS 모듈은 텍스트를 음성으로 변환하는 기능을 검증합니다. 다양한 텍스트를 입력으로 제공하여, 변환된 음성이 자연스러운지 확인합니다.

<통합 테스트>

모든 모듈이 서로 잘 연동되어 작동하는지 확인합니다. 이를 위해 시스템 전체를 통틀어 테스트하며, Use Case가 의도한 대로 작동하는지 확인합니다.

Use Case (사용자 감정 인식 및 대화 생성)

1. 사용자가 카메라 앞에 위치합니다.
2. Camera 모듈이 사용자의 얼굴 이미지를 캡처합니다.
3. Frame 모듈이 캡처된 이미지를 처리합니다.
4. Face Detection 모듈이 처리된 이미지에서 얼굴을 탐지합니다.
5. Crop 모듈이 탐지된 얼굴 부분을 잘라냅니다.
6. Emotion 모듈이 잘라낸 얼굴 이미지를 분석하여 사용자의 감정 상태를 판단합니다.
7. Chat 모듈이 판단된 감정 상태를 바탕으로 대화 스크립트를 선택합니다.
8. TTS 모듈이 생성된 대화를 음성으로 변환하여 사용자에게 전달합니다.
9. Microphone 모듈이 사용자의 음성 입력을 받아 텍스트로 변환합니다.
10. ChatGPT 모듈이 변환된 텍스트를 바탕으로 적절한 대화를 생성합니다.
11. TTS 모듈이 생성된 대화를 음성으로 변환하여 사용자에게 전달합니다.

<성능 테스트>

웹캠을 통해 받은 영상에서 얼굴을 인식하고 감정이 나오기까지 걸리는 한 프레임의 시간을 측정합니다.