**TORRENS UNIVERSITY AUSTRALIA**

**Applications of Artificial Intelligence AAI202**

**Assessment 3**

**Natural Language Processing and Computer Vision Problem Sets**

**Code included in ZIP folder with datasets**

**Luis Alberto Mercado Rodriguez**

**Student ID A00147277**

[**LUIS.RODRIGUEZ1@STUDENT.TORRENS.EDU.AU**](mailto:LUIS.RODRIGUEZ1@STUDENT.TORRENS.EDU.AU)

**2024**

**Abstract**   
This report explores the development of two artificial intelligence systems: speech recognition and face detection. The speech recognition system converts spoken language into text, utilizing Python's speech\_recognition library and Google's speech recognition API. This system works by listening to audio inputs through a microphone, adjusting for background noise, and processing the audio to convert it into text. The face detection system identifies and recognizes human faces in live videos, employing the face\_recognition library and OpenCV. The system uses a dataset of known faces to create unique numerical codes, called face encodings, which are then compared to detected faces in live videos to identify matches. Testing and validation of both systems with real data produced impressive results, demonstrating their effectiveness and reliability in various scenarios.

**Introduction**

Artificial intelligence (AI) has transformed interactions between humans and machines, enabling computers to understand and simulate human abilities such as **Speech recognition**, machines are now able to convert spoken words into text. Making it possible now to create applications that can understand human speech and perform actions according to it. On the other hand, it is also possible to detect faces a computer can identify human faces in images and videos creating multiple possibilities of implementation, from pattern recognitions that can be used in security cameras to simple image classifiers. This report explores the development and testing of these AI systems, explaining their capabilities, and approaches in how to use them and validate them.

**Speech Recognition**

Speech recognition converts spoken language into text. For this project, we used Python's speech\_recognition library. The system works by listening through a microphone and adjusting for background noise. When it hears speech, it processes the audio and converts it to text using Google's speech recognition API. (*Speech Recognition Examples With Python - Python*, n.d.)

A diagram of a diagram

Description automatically generatedThe main part of the system is a class that manages the microphone and listening process. When it recognizes speech, it prints the text. If it can't understand, it keeps listening.

**Testing and Validation**

We tested the system with live speech through a microphone and pre-recorded audio files. The pre-recorded files had known sentences, which helped us check the accuracy of the system. For this, we used two files retrieved from Kaggle (*Sample Audio Files for Speech Recognition*, 2020), one with clear audio but difficult words and multiple sentences and the second one with a noisy background and just one sentence.

Because is normal for this system to make mistakes for edgy words, instead of just checking if the whole text matches the expected text, a word-by-word approach was implemented. So it gives a percentage of accuracy based on how many words are correct. For the first audio, an accuracy of 93.02% was received with 40 correct words out of 43 and for the second audio with noisy background, the accuracy was 85.71% with 6 out of 7 words correct.

C:\...\known\_speech\harvard.wav:

RECOGNIZED TEXT: the stale smell of old beer lingers it takes heat to bring out the o'dor a cold dip restores health and zest a salt pickled tastes fine with ham tacos Al pastora are my favorite a zestful food is the hot cross bun

CORRECT TEXT: the stale smell of old beer lingers it takes heat to bring out the o'dor a cold dip restores health and zest a salt pickle tastes fine with ham tacos al pastor are my favorite a zestful food is the hot cross bun.

Accuracy: 93.02% (40/43 correct words)

C:\...\known\_speech\jackhammer.wav:

RECOGNIZED TEXT: the stairs smell of old beer lingers

CORRECT TEXT: the stale smell of old beer lingers

Accuracy: 85.71% (6/7 correct words)

In live tests, the system accurately transcribed spoken sentences. This speech recognition system is effective and reliable for capturing and transcribing spoken language.

**Face Detection**

Face detection involves identifying and recognizing human faces from live videos recorded by the camera. The process starts by preparing the data images of known people, for this report, a dataset of 14 celebrity faces from Kaggle (*14 Celebrity Faces Dataset*, 2019) was used, which is stored in folders named after each person for example, the faces of Luis Mercado are in a folder named luis\_mercado inside the known\_faces folder, we do this to associate images to each person. During training, the system scans these images, detects faces, and converts them into unique numerical codes called face encodings using the face\_recognition library. These encodings A diagram of a face detection process

Description automatically generatedcapture key facial features and are saved for future comparison and testing.

The system uses two main face detection models: HOG (Histogram of Oriented Gradients) and CNN (Convolutional Neural Network) (Bilal\_Ai, 2023). HOG is fast and works well for real-time tasks, but CNN offers better accuracy, especially in complex scenarios where the images are not clear enough or there is not much relationship between both images. It works in the way that when the system detects a face, it compares its encoding with the saved ones. If the match or distance is close enough (in this case below 0.6), the face is recognized; otherwise, it's labeled "Unknown." The results are displayed on a live video, with detected faces marked in a square and named using OpenCV.

**Testing Results:**

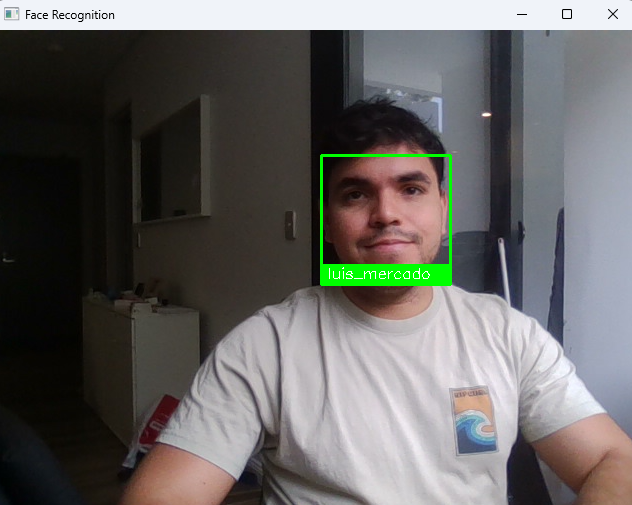
The system was tested in two ways, the first one is the live webcam detection and validation using new images. In the webcam test, it successfully identified known faces and marked unfamiliar ones as "Unknown," proving its real-time accuracy.

During validation with unseen images, the system achieved 100% accuracy. This means it correctly identified faces in 100 out of 100 cases. The precision (100%) indicates that it correctly labeled all the detected faces, while the recall (100%) shows it identified most of the faces it should have. These results help us to know the reliability of the system and how efficient is with the face recognition.

A screen shot of a black background

Description automatically generated

**Testing Outputs of 3 users.**

**Luis Mercado**

**A hand holding a cell phone

Description automatically generatedBen Afflek**

**A person and person with their faces in a photo

Description automatically generatedSofia Vergara**

**Conclusion**

The face detection and speech recognition technologies implemented in this report have demonstrated a consistent performance. Spoken words were accurately translated into text by the speech recognition technology. Faces in live recordings were effectively recognized by the face detection system and correctly labeled. There are numerous potential applications for these systems, and clearly, they need improvements to enhance their effectiveness and accuracy. But this is a good point to start learning about these systems and what is behind them, how they work, and how to approach further applications.

**References**

*14 Celebrity Faces Dataset*. (2019, August 19). <https://www.kaggle.com/datasets/danupnelson/14-celebrity-faces-dataset/data>

*sample audio files for speech recognition*. (2020, August 14). <https://www.kaggle.com/datasets/pavanelisetty/sample-audio-files-for-speech-recognition/data?select=jackhammer.wav>

*Speech Recognition examples with Python - Python*. (n.d.). <https://pythonprogramminglanguage.com/speech-recognition/>

Bilal\_Ai. (2023, October 6). HOG vs. CNN “Unveiling the Power of Image Classification.” *Medium*. <https://medium.com/@entrepreneurbilal10/hog-vs-cnn-unveiling-the-power-of-image-classification-a3d1585d76e8>

**CODE INCLUDED IN THE ZIP FOLDER WITH DATASETS**