Implementation of an Audio-Visual Emotional Recognition

Merve DADAŞ, Furkan KARADAŞ, Aydın ŞİŞMAN, Uğur BAYEZİT
Assoc. Prof. Dr. Hadi Hakan MARAŞ & Assist. Prof. Dr. Selma ÖZAYDIN

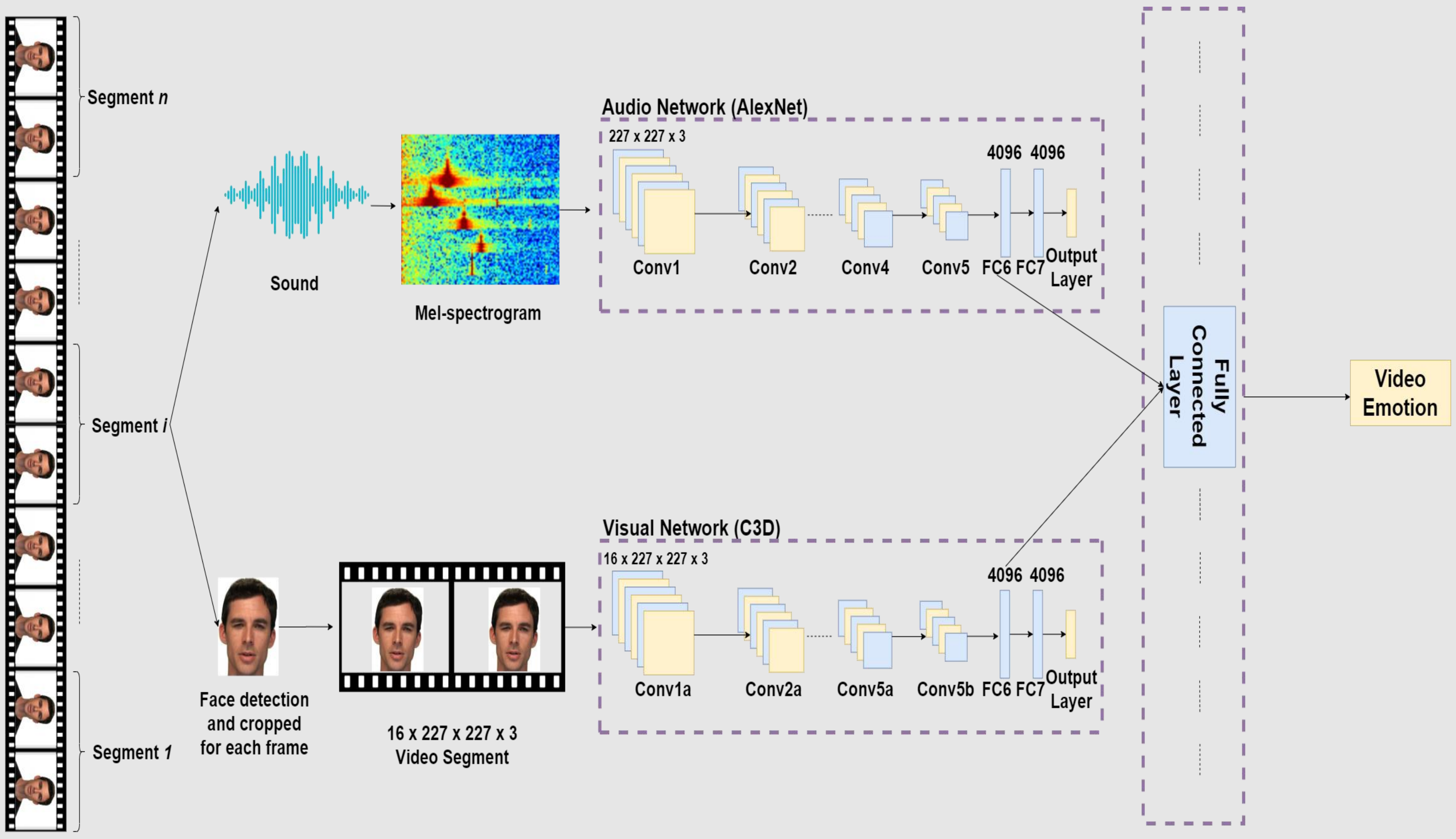
Çankaya University, Department of Computer Engineering

Abstract

People express emotions through differently. Utilization of both verbal and nonverbal communication channels allows to create a system in which the emotional state is expressed more clearly and therefore easier to understand. In this report describes the emotion recognition project consisting of audio and visual. Emotion recognition will be performed with features extracted from speech and video. In the development of the project, image processing, signal processing and artificial intelligence technologies were used. The audio and video were processed separately and then combined with the fusion algorithm.

Key words: Image processing, speech processing, signal processing, deep learning, machine learning, classification

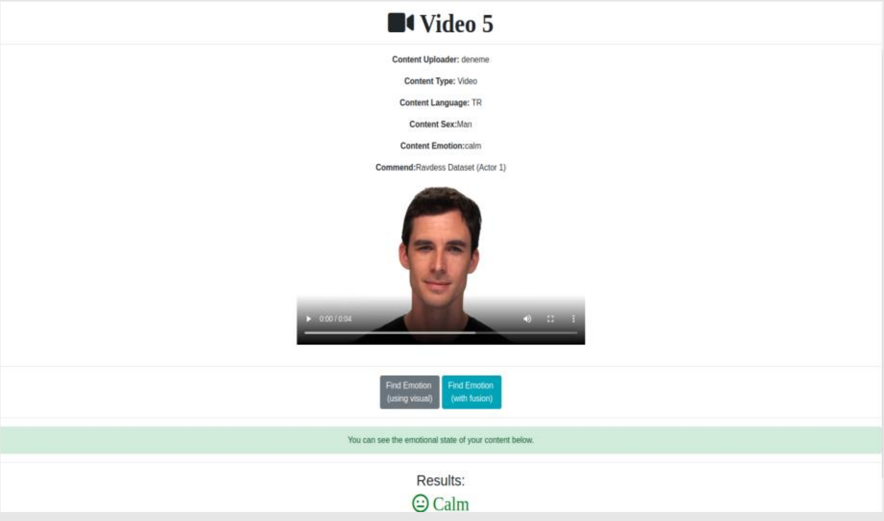
Algorithms



Introduction

Modern day security systems rely heavily on bio-informatics, like as speech, fingerprint, facial images and so on. Besides, determination of a user's emotional state with facial and voice analysis plays a fundamental part in man-machine interaction (MMI) systems, since it employs non-verbal cues to estimate the user's emotional state. Therefore, recognizing human emotion has been an attractive task for data scientists. On the other hand, there are many challenges in emotional data evaluation such as collection of proper datasets, definition of number of emotions to recognize, selection of the labelled data, etc. Due to the many challenging tasks under evaluation, MMI systems that utilize multimodal information about their users' current emotional state are interest of the computer vision and artificial intelligence communities.

Piece of Product



Video 5

Content Uploader: deneme

Content Type: Video

Content Language: TR

Content Sex: Man

Content Emotion:calm

Commend: Ravdess Dataset (Actor 1)

You can see the emotional state of your content below.

Results:

☺ Calm

Techniques Used

During the implementation stage, transfer learning and deep learning procedures have been used for feature extraction of videos. "RAVDESS SONG" and "RAVDESS SPEECH" datasets were used to train the system. These techniques are that below:

Audio Part

- **Machine Learning Techniques**
  - Multi-Layer Perception Classifier
  - Linear Discriminant Classifier
  - K-Nearest Neighborhood
  - Support Vector Classifier
- **Deep Learning Technique**
  - **Preprocessing**
    - Split frame for transfer learning
  - **Transfer Learning**
    - AlexNet Model

Visual Part

- **Preprocessing**
  - Face Detection & Crop
  - Resize
- **Deep Learning Techniques**
  - **Transfer Learning**
    - C3D Model
  - **Deep Feature Extraction**
    - Support Vector Machine
    - K-Nearest Neigborhood
    - Random Forest
    - Decision Tree

Fusion Part

- Fully Connected Layer

GUI Part

- Django Framework

Conclusion

The audio and visual emotion recognition system has been successfully completed. When you upload an audio-visual video, audio and video are separated and processed, and then recombined and passed through a special algorithm, giving you the most accurate emotion.

Acknowledgement

We are grateful for guidance we have received from **Assoc. Prof. Dr. Hadi Hakan MARAŞ** and **Assist. Prof. Dr. Selma ÖZAYDIN**. The help we received from them was a great asset to improve this project and ourselves. Also we would like to thanks all CENG departments lecturers, especially **Assist. Prof. Dr. Nurdan SARAN**.

Group Members

Quarantine days meeting due to Covid-19