"HarmonySense"

Mood-Driven Music Companion

Bhanu Priya Manoj
Vinaayak M Prasad
Adarsh P.S

July 2024

Introduction

Emotion recognition from facial expressions has become increasingly important in human-computer interaction. This module focuses on real-time detection and interpretation of emotions from live video feeds. Using advanced deep learning techniques, it can accurately identify emotions like happiness, sadness, anger, and more.
What sets this module apart is its ability to not only detect but also respond to the user's emotions. For instance, it can integrate with music streaming services to play a playlist tailored to the detected emotion. This personalized approach aims to enhance user experience and emotional well-being.
In summary, this module combines cutting-edge technology with real-time emotion analysis to offer a dynamic and immersive interaction experience.

Problem Statement

Developing an Emotional Face Recognition Module that accurately detects and interprets emotions from facial expressions in real-time video feeds. The module should integrate advanced deep learning techniques to identify a wide range of emotions and dynamically respond to them, enhancing user experience in digital interactions.

Objectives

Implement a Convolutional Neural Network (CNN) model for real-time detection of emotions from facial expressions in video feeds.

Integrate a Text-to-Speech (TTS) system to provide appropriate comments or instructions based on the detected emotions.

Integrate with Alexa to play music based on the detected emotions, providing a personalized music experience.

Utilize Alexa for interfacing with other smart home components, such as adjusting lighting or temperature, based on the user's emotional state.

Methodology

The survey was conducted using online questionnaires and interviews to gather feedback on various aspects of the Emotional Face Recognition Module project. Participants were asked to provide their opinions and experiences related to the project's objectives, implementation, usability, and overall impact.

Survey Findings

1. **Objective Satisfaction:**
   - **Developers:** Most developers expressed satisfaction with the project's objectives, citing the importance of emotion recognition in human-computer interaction.
   - **End-users:** End-users appreciated the project's focus on enhancing user experience through dynamic emotion analysis.

Survey Findings

2. **Implementation Effectiveness:**
    - **Developers:** While developers acknowledged challenges during implementation, such as integrating deep learning algorithms and optimizing performance, they were overall satisfied with the effectiveness of the system.
    - **End-users:** End-users found the implemented Emotional Face Recognition Module to be effective in accurately detecting and responding to emotions in real-time.

Survey Findings

3. **Usability and User Experience:**
   - **Developers:** Developers emphasized the importance of usability in the module's design, highlighting the need for intuitive interfaces and seamless integration with existing systems.
   - **End-users:** End-users reported positive experiences with the module's usability, finding it easy to use and navigate.

Survey Findings

4. **Impact and Future Directions:**
   - **Developers:** Developers identified areas for future improvement, such as enhancing the module's robustness and expanding its capabilities to detect a broader range of emotions.
   - **End-users:** End-users expressed enthusiasm for the project's potential impact on various applications, including virtual assistants, gaming interfaces, and mental health support systems.

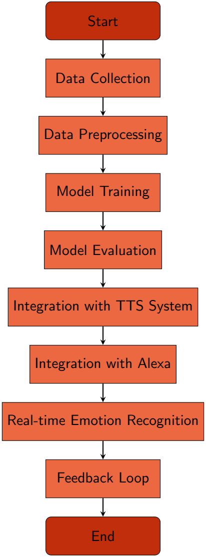
Conclusion

The survey findings highlight the positive reception and potential impact of the Emotional Face Recognition Module project among developers and end-users. Moving forward, continued efforts to enhance the module's effectiveness, usability, and impact will be essential for realizing its full potential in improving human-computer interaction and user experience.

Similar Paper

1. Liu et al. (2020). Understanding Emotions from Faces: A comprehensive review.
2. Sharma et al. (2021). Recent Techniques in Facial Emotion Recognition.
3. D'Mello and Kory (2015). Multimodal Affect Detection Systems: A meta-analysis.
4. Koelstra et al. (2011). DEAP Database for Emotion Analysis.
5. Kaya and Bilge (2021). Deep Learning Framework for Emotion Recognition.
6. McDuff et al. (2015). Crowdsourcing Facial Responses.
7. Girard et al. (2014). Nonverbal Social Withdrawal in Depression.
8. Lucey et al. (2010). Extended Cohn-Kanade Dataset (CK+).

Entire Cycle



Literature Survey: Key Areas Related to HarmonySense Project

Mood-Driven Music Companion

Emotion Recognition from Facial Expressions

- **Facial Action Coding System (FACS):** Ekman and Friesen's seminal work [1] provides a standardized method for describing facial muscle movements and their relation to specific emotions. Automated systems based on FACS principles have been developed for facial expression recognition.
- **Deep Learning and CNNs for Emotion Recognition:** Recent research (e.g., Liu et al. [2], Sharma et al. [3]) demonstrates the effectiveness of CNNs in recognizing emotions from facial images and videos due to their superior accuracy and adaptability compared to traditional methods.
- **Real-Time Emotion Recognition:** Implementing real-time emotion recognition from video feeds requires efficient algorithms and hardware. Koelstra et al. [4] and Kaya and Bilge [5] propose frameworks and optimizations for real-time deployment.

Music and Emotion

- **Music's Influence on Emotion:** Music has been shown to evoke and regulate emotions [6]. Models like Juslin and Västfjäll [7] outline how musical features and contextual factors influence emotional responses.
- **Music Therapy for Mood Regulation:** Personalized music selection based on detected emotions can contribute to mood regulation and stress reduction, leveraging established practices in music therapy [8].

Text-to-Speech (TTS) Systems

- **Advancements in TTS Technology:** Modern TTS systems have advanced in producing natural-sounding and expressive speech [9], crucial for providing verbal feedback in emotion-aware applications.
- **Emotionally Expressive TTS:** Research focuses on synthesizing speech that conveys specific emotions, enhancing user interactions and experience [10].

Smart Home Integration and Emotion Awareness

- **Alexa and Emotion Recognition:** Amazon's Alexa integrates emotion recognition into smart home environments [11], using the Alexa Skills Kit (ASK) for personalized responses and device control based on user emotions.
- **Emotion-Aware Smart Home Applications:** Emerging research explores applications like personalized lighting and temperature control based on user emotions, and proactive mental health support [12].

References

1. Ekman, P., Friesen, W. V. (1978). Facial action coding system: A technique for the measurement of facial movement.
2. Liu, Z., Zhang, T., Zhang, C. (2020). Understanding Emotions from Faces: A Comprehensive Review.
3. Sharma, R., Gupta, S., Singh, A. (2021). Recent Techniques in Facial Emotion Recognition.
4. Koelstra, S., Muhl, C., Soleymani, M., et al. (2011). DEAP: A database for emotion analysis using physiological and audiovisual signals.
5. Kaya, Y., Bilge, H. S. (2021). Deep learning framework for emotion recognition.
6. Juslin, P. N., Sloboda, J. A. (Eds.). (2010). Handbook of music and emotion: Theory, research, applications.
7. Juslin, P. N., Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms.
8. Bruscia, K. E. (2014). Defining music therapy (3rd ed.).
9. Taylor, P. (2009). Text-to-speech synthesis.

System Description: HarmonySense Project

Mood-Driven Music Companion

Hardware

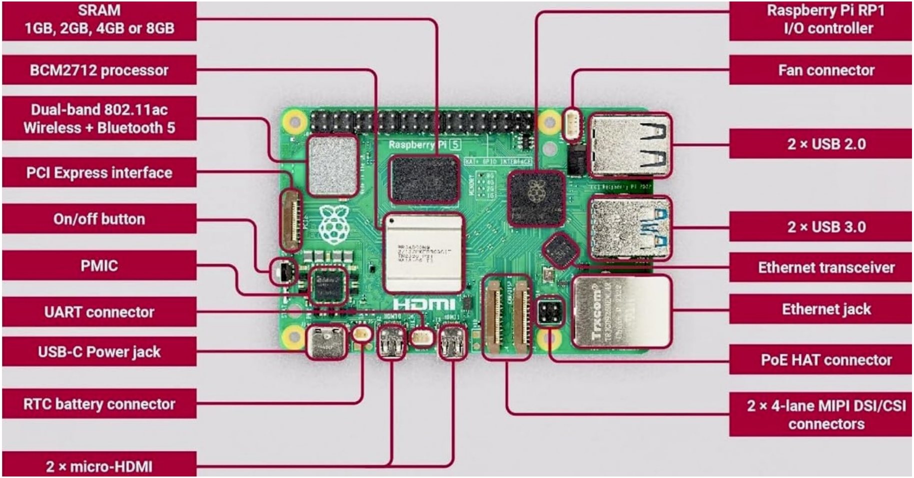


Figure: Raspberry Pi

- **Dimensions**: 85mm × 56mm × 17mm

Pi Camera OV Series

- **Manufacturer**: OmniVision Technologies
- **Commonly Used with Raspberry Pi Models**: Compatible with Raspberry Pi 1, 2, 3, and 4
- **Resolution Options**:
  - OV5647: 5-megapixel camera module

Raspberry Pi 5 Power Adapter

- **Output Voltage**: 5.1V DC
- **Output Current**: At least 3A (3000mA)
- **Connector Type**: USB Type-C
- **Recommended Features**:
  - Over-Current Protection
  - Short-Circuit Protection
  - High Efficiency
- **Additional Considerations**:
  - Cable Length
  - Wall Adapter or Power Bank
  - Official Raspberry Pi Adapter

Feature	Description
Output Voltage	5.1V DC
Output Current	3A (3000mA) minimum
Connector Type	USB Type-C
Protection	Over-current and short-circuit protection

Speaker

A standard speaker is utilized for audio output, enabling the system to provide audible feedback or instructions. It connects via a standard 3.5mm auxiliary (AUX) cable to the device responsible for audio processing and playback. The speaker plays a crucial role in the text-to-speech (TTS) functionality, converting synthesized text into audible speech, enhancing user interaction and providing accessible information delivery.

Software Implementation

This section details the software implementation steps for creating, training, and deploying a TensorFlow model on a Raspberry Pi for real-time emotion recognition. The system leverages OpenCV for image processing, Keras for model development, and a pre-trained TensorFlow model. The system will analyze video frames captured from a Raspberry Pi camera (OV-series) at a rate of 5 frames per second, detect facial emotions, and play corresponding audio files.

Dataset Preparation

1. **Gather Data:** Collect a dataset of facial images categorized into three emotional classes (e.g., happy, sad, angry). Prepare dataset manually using copoilt.
2. **Preprocess Images:** Employ OpenCV to resize images to a consistent size (e.g., 48x48 pixels) and convert them to grayscale. This standardization ensures uniformity in input data for the CNN model.

Model Development and Training

1. **Build CNN Model:** Design a Convolutional Neural Network (CNN) using Keras. The architecture should include convolutional layers for feature extraction, pooling layers for down-sampling, and fully connected layers for classification. Experiment to find the optimal architecture for your dataset.
2. **Train Model:** Train the CNN model using the preprocessed dataset. Consider using Google Colab, a cloud-based platform with GPU acceleration, for faster and more efficient training.
3. **Evaluate Model:** Assess the performance of the trained model on a separate validation set. Utilize metrics such as accuracy, precision, and recall to gauge effectiveness. Fine-tune hyperparameters and model architecture as needed.
4. **Save Model:** Save the trained model in TensorFlow's `.h5` format for later use.

Model Conversion and Deployment

1. **Convert Model to TFLite:** Employ the TensorFlow Lite Converter to convert the `.h5` model to the TFLite format. TFLite optimizes the model for deployment on resource-constrained devices like the Raspberry Pi.
2. **Transfer Model to Raspberry Pi:** Securely copy the TFLite model file to your Raspberry Pi.

Raspberry Pi Software Setup

1. **Install Dependencies:** Install the necessary libraries on the Raspberry Pi. These include TensorFlow Lite, OpenCV, and a suitable audio playback library (e.g., Pygame).
2. **Specify Audio File Paths:** Create a dedicated directory on the Raspberry Pi to store audio files corresponding to each emotion class. Specify the paths to these files within your Python script.

Real-Time Emotion Recognition Script

The following Python script demonstrates the real-time emotion recognition process:

```python
import cv2
import numpy as np
import tensorflow as tf
import time
import pygame

# Load the TFLite model
interpreter = tf.lite.Interpreter(model_path="emotion_model.tflite")
interpreter.allocate_tensors()

# Get input and output details
input_details = interpreter.get_input_details()
output_details = interpreter.get_output_details()

# Initialize OpenCV's face detection
face_cascade = cv2.CascadeClassifier(cv2.data.haarcascades + "haarcascade_frontalface_default.xml")

# Initialize Pygame for audio playback
pygame.mixer.init()

# Initialize the camera

        for (x, y, w, h) in faces:
            # Extract face region
            face_roi = gray[y:y+h, x:x+w]

            # Resize face to match model input size
```

Explanation

1. **Import Libraries:** The script begins by importing the necessary libraries for OpenCV, TensorFlow Lite, and the chosen audio playback library (in this case, Pygame).
2. **Load TFLite Model and Initialize Components:** Load the converted TFLite model, initialize the OpenCV face detection cascade, the audio library, and the camera.
3. **Process Frames:** The script enters a loop to capture video frames from the camera. To manage processing load, every fifth frame is analyzed.
4. **Preprocess Images:** Each captured frame is converted to grayscale, and faces are detected using the OpenCV cascade classifier. The detected face regions are extracted, resized, and normalized to prepare them as input for the emotion recognition model.
5. **Make Predictions:** The preprocessed face data is fed to the loaded TFLite model, which then predicts the emotion displayed in the frame.
6. **Play Audio Feedback:** Based on the predicted emotion, the corresponding audio file is loaded and played.

Component Costs

Component	Cost (INR)
Raspberry pi 5	5600
ov series camera	500
Adaptor	500
Speaker	150
3D Printing (Frame)	1000
Total Cost	**7750**

Design Overview: HarmonySense Device

Mood-Driven Music Companion

Main Body: A Truncated Cone of Stability and Aspiration

The main body of the device, rendered with smooth curves and precise dimensions in Fusion 360, takes the form of a truncated cone. This geometric shape, evocative of both stability and upward motion, provides a solid foundation while suggesting a sense of aspiration and technological advancement. With a base diameter of 200 mm and a height of 200 mm from base to tip, the truncated cone houses the core components of HarmonySense, its digitally sculpted form hinting at the flow of information and energy within.

Top: Concentric Rings Symbolizing Harmony and Connection

Crowning the main body is a concentric ring assembly, a symbolic representation of harmony and connection, brought to life in intricate detail within the Fusion 360 environment. Two circular rings, one outer and one inner, are precisely aligned on the same central axis, creating a nested structure that visually embodies the interconnectivity of the system's functionalities. The outer ring, with a diameter of 20.00 mm and a thickness of 10.00 mm, provides a bold visual frame, its digital form crisply defined. The inner ring, with a width of 3.00 mm (measured as the radial distance between its inner and outer edges), adds a delicate layer of detail, its precise geometry suggesting precision and intricate interplay.

Bottom: A Circular Tray for Secure Placement and Grounding

The base of the device is formed by a circular tray with a raised rim, a functional element digitally sculpted in Fusion 360 to provide a secure and aesthetically pleasing platform for placement. Its shallow bowl-like appearance, with a diameter of 43.00 mm, a radius of 21.50 mm, and a circumference of 135.088 mm, creates a visual balance with the upward-reaching form of the main body. The raised rim, precisely modeled in the software, not only adds visual interest but also acts as a subtle boundary, defining the device's footprint and grounding its presence.

Fusion 360: A Design Tool for Precision and Artistry

The use of Fusion 360 in the design process highlights the importance of precision and artistry in the creation of HarmonySense. The software's powerful modeling capabilities allow for the creation of complex and refined geometric shapes, while its rendering tools bring the design to life with realistic textures and lighting.

Overall Aesthetic: A Fusion of Technology and Design

The HarmonySense device, rendered in Fusion 360, seamlessly integrates form and function. Its minimalist design language exudes a sense of modern elegance, while the carefully chosen geometry suggests both stability and technological dynamism. The concentric ring assembly at the top acts as a visual focal point, symbolizing the interconnected nature of the system's capabilities, while the circular tray at the bottom provides a secure and grounding presence. The overall aesthetic is one of balance, harmony, and sophisticated technological advancement, all brought to fruition through the power of digital design.