

AI-Based Camera: Integrated Speech Recognition and Emotion Detection



NORTHERN ILLINOIS UNIVERSITY

College of Engineering and Engineering Technology

K. Anupama Satya Sai Lakshmi¹, Keerthi A. M.¹, Prasanthi V¹, Sai Dhruthi Varna K¹, K. Sudarsana Reddy²
Dr. Mansour Tahernezehadi³

¹Electronics and Communications Engineering, ²Electrical and Electronics Engineering and ³Electrical Engineering

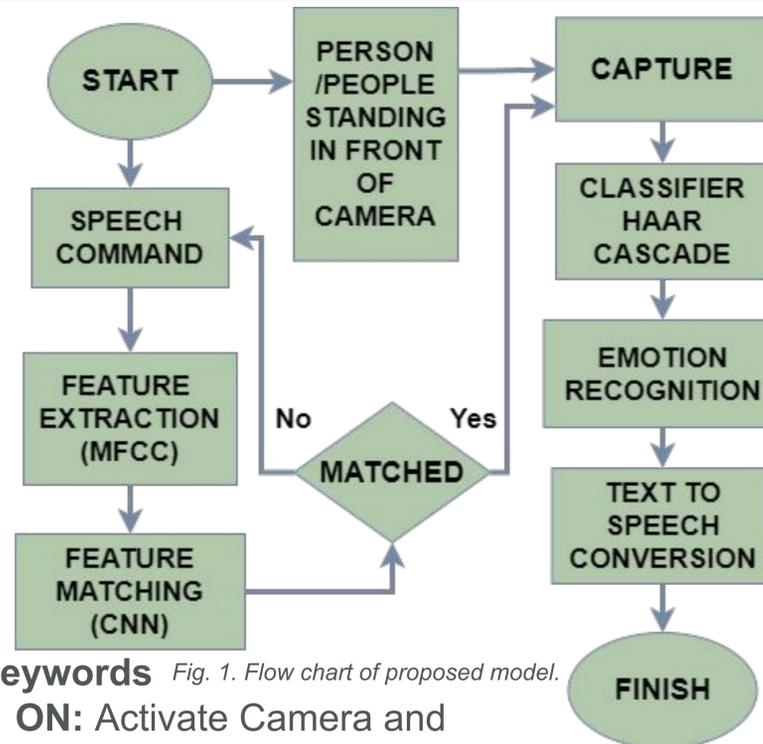
Abstract

In the current scenario, Artificial Intelligence is emphasizing the development of intelligent machines. Considering the frequent use of web camera applications, there is a need for an enhanced and automated version of the application. Emotion Recognition also plays a crucial role in daily lives. An integrated solution of speech detection with the camera application is a great duo and handy to the user. This project presents a method for accessing the camera to take an image and perform image processing using pre-trained voice commands. Additionally, the model also predicts the emotion of the user and produces speech output of the emotion recognised.

Problem Statement

The world's growing demand requires simultaneous advancements in automation. One such required demand is a need for automated hands-free camera application. The focus of the project is to develop a user-friendly camera application using pre-defined keywords by applying different post-image processing techniques and detecting emotion as per the user's requirement. The main challenge is to integrate speech and face recognition as one system.

Methods



Keywords Fig. 1. Flow chart of proposed model.

- **ON:** Activate Camera and Click picture
- **ONE:** RGB Filter
- **TWO:** B&W Filter
- **THREE:** Auto Adjustment of Brightness and Contrast
- **FOUR:** Zoom-in
- **FIVE:** Zoom-out
- **SIX:** Emotion Recognition

Analysis

ANALYSIS OF EMOTION RECOGNITION MODEL

EMOTION	TOTAL TRAIN SAMPLES	TOTAL TESTING SAMPLE	NUMBER OF CORRECTLY CLASSIFIED IMAGES	SENSITIVITY
HAPPY	8090	899	824	0.92
SAD	5469	608	400	0.66
NEUTRAL	5578	620	446	0.72

Results and Discussion

According to the keywords identified by the speech recognition model, the post image processing techniques are applied to the captured image and the emotion can also be recognized. The accuracy of the speech recognition model and emotion detection model is 94% and 78.7% respectively.



Fig. 2. (a) Captured image, (b) RGB filter output, (c) B&W filter output.

Conclusions

A camera application that operates on speech commands was developed. Automatic speech recognition helps to click images and apply five post-processing image techniques and emotion recognition. The model recognizes three different emotions namely happy, sad and neutral. This model can also be used in a variety of real-time applications. Future work can focus on increasing the accuracy of the emotion detection model.

Acknowledgements

We would like to express our gratitude to Dr. Mansour Tahernezehadi from the department of electrical engineering for his technical and moral support throughout the project.