Video Analysis With Speech Modulation Using Neural Networks

[1]Jaymala Pawar, [1]Shamal Pawar, [1]Mrunal Pardeshi, [1]Abhishek Patil

[2]*Prof.M.P.Nerkar*

[1]*Research Scholar, Computer Department, AISMMS IOIT, Pune*

[2]*Professor, Computer Department, AISMMS IOIT, Pune*

Abstract

Many devices now-a-days are having a capacity to interact with the end users. It is a challenge for these devices to understand analyse the sentiments of a user. Real-time analysis of speech patterns and facial expressions for detecting sentiments. Many deep learning techniques prove useful in such cases. Our literature survey well explains facial expression analysis and speech to text conversion methods. This paper explains semantic localization technique. Semantic networks have been used for making sentence hypotheses for the purpose of speech recognition. Natural Language Processing (NLP) plays an important role to compute the semantics between two short texts. It is challenging enough to give high accuracy output using feature extraction and classification schemes. Hence we aim to achieve high performance solution in order to be efficiently implemented on hardware like GPU(NVIDIA). This paper specifies the ideology and the methodology that facilitates the user to evaluate its sureness and accuracy for a certain article.

Keywords *Image Processing, Semantics, Neural Networks, Modulation, Cloud Computing, Data Mining, Facial Expressions , Key Frames.*

Introduction

Because of its various potential applications including psychology, medicine, security, man-machine interaction ,surveillance, interview process and for physically disabled people, in recent years the facial expressions classification has attracted a lot of attention. Action Unit (AU) based and appearance-based methods are the two main approaches to investigate the facial expression . This technique described the facial expression as a composition of Action Units which are describing the facial muscle motions. Since it uses the facial muscle movements for modelling different expressions this method takes advantage of the strong support of the psychology and physiology sciences. It is extremely difficult to model the system using machines due to the AU based methods suffering from the difficulties such as dependencies on invisible muscle motions .To detect image expressions and YOLO libraries for real time object detection, here the idea is to use Google Vision API.
Having limited vocabularies of about a dozen words earlier systems were limited to a single speaker. Having enormous vocabularies in numerous languages, modern speech recognition systems can recognize speech from multiple speakers. Of course, speech is the first component of speech recognition. Speech must be converted from physical sound to an electrical signal with a microphone, and then to digital data with an analog-to-digital converter. To transcribe the audio to text several models can be used. Using techniques for feature transformation in many modern speech recognition systems, neural networks are used to simplify the speech signa. Hence, we combine both the approaches i.e. Facial expression analysis and speech to text conversion to analyse the efficiency and accuracy of the speech and mood or emotional state of the user so that this module can be used as part in many applications as mentioned in the later part.

Literature Survey

1]In the paper "Discourse To-Text transformation progressively" proposed by Nuzhat Atiqua Nafis and Md. Safaet Hossain they build up a product that upgrades the client's method for discourse through accuracy of articulation following the English phonetics. This product permits one to learn, pass judgment and perceive their potential in English language. This frameworks has applications in intelligent voice reaction framework (IVRS) ,voice-dialing in cell phones and phones ,without hands dialing in remote bluetooth headsets Speech to Text Conversion in Real-time , PIN and numeric secret phrase section modules , computerized teller machines (ATMs) , information passage work , in homeroom works for impaired understudies.

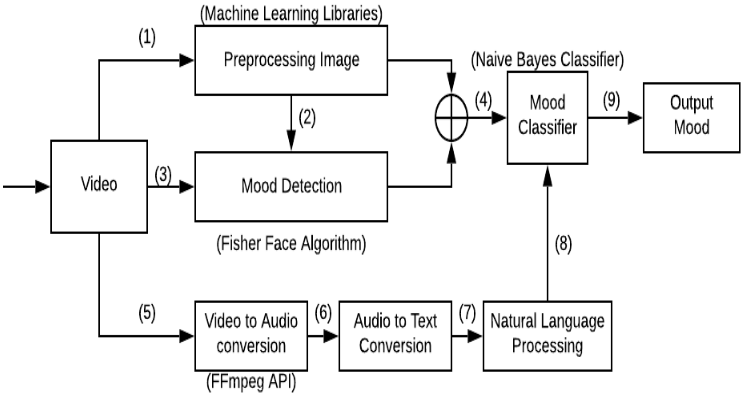
2] To measure the semantic similarity between academic papers and patents, using three methods like the Jaccard coefficient, cosine similarity of tf-idf vector, and cosine similarity of log-tf-idf vectors, N. Shibata et. al performed a comparative study. They also performed a case-study for further analysis and all of these methods are corpus-based methods. S. Zhang et. al explored the different applications of semantic similarity related to the field of Social Network Analysis

3]H. Pu et. al incorporated a semantic similarity calculation algorithm that used the large corpus to compare and analyse several words. This method used the Word2Vec model to calculate semantic information of the corpus and used Li similarity measure to compute similarity. Wael H. Gomaa also discussed three text similarity approaches: corpus-based, string-based and knowledge-based similarities and proposed different algorithms based on it.

4] Sakai et al presented Face identification dependent on edges This work depended on dissecting line drawings of the appearances from photos, planning to find facial highlights. Than later Craw et al. proposed a various leveled system dependent on Sakai et al's. work to follow a human head layout. A technique proposed by Devaranjan and Anila was basic and quick. They proposed structure which comprise three stages for example , at first the pictures are upgraded by applying middle channel for clamor evacuation and histogram evening out for differentiate change. In the second step the edge pictures built from the upgraded picture by applying sobel administrator. To separate the sub windows from the improved picture dependent on edges a novel edge following calculation is applied. To arrange the sub-window as either face or non-face further they utilized Back engendering Neural Network (BPN) calculation.

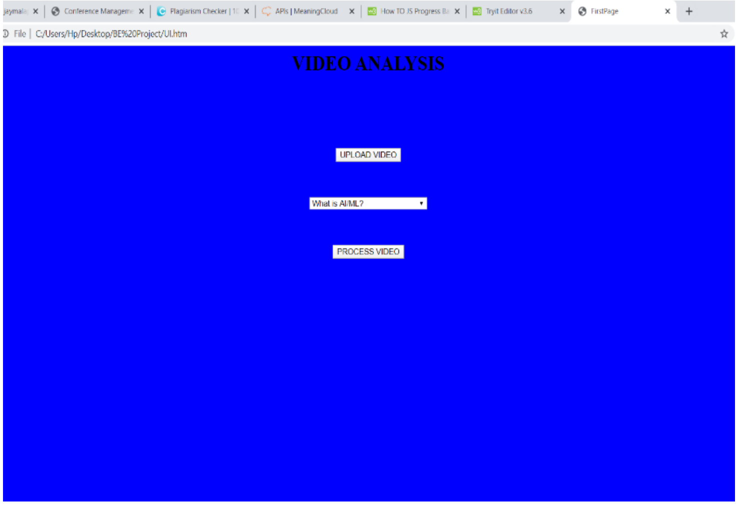
5]In the paper for face recognition using Fisherface method the research is done to establish a program of face recognition application by utilizing GUI applications and databases.

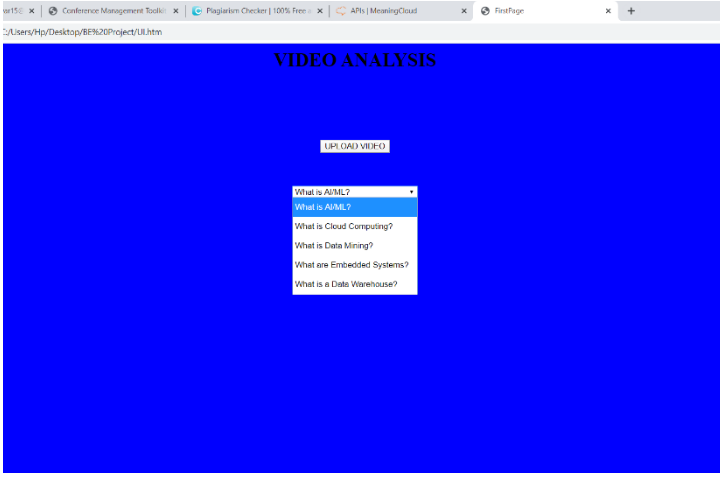
System Architecture



Proposed Methodology:

1.First of all, we are taking the video as input and convert it into audio. For this we have used FFmpeg library. It is very fast video and audio converter that can also grab from live audio/video source.




ar15@
Conference Management Toolkit
Plagiarism Checker | 100% Free a
APIs | MeaningCloud
FirstPage
C:/Users/Hp/Desktop/BE%20Project/UI.htm
VIDEO ANALYSIS
UPLOAD VIDEO
What is AI/ML?
What is AI/ML?
What is Cloud Computing?
What is Data Mining?
What are Embedded Systems?
What is a Data Warehouse?

2.After the conversion of video to audio, next step is to convert it into the text and for this we have used SpeechRecognition API. This API is used to understand the words that are being spoken by human being.

3.After the audio to text conversion, for checking the efficiency of the text semantically, TF-IDF Algorithm is used.

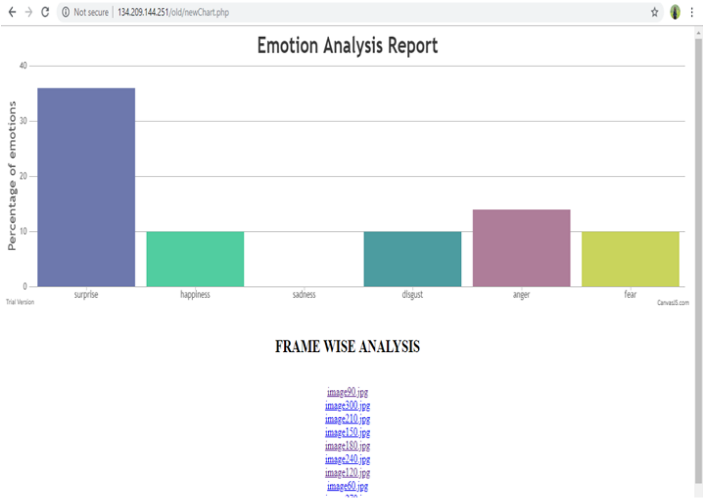
4.Then for the extraction of the frames from the video, OpenCV is used. It is a cross-platform library using which we can develop real-time computer vision application. It mainly focuses on image processing, video capture and analysis including features like- Face detection and Object detection.

5.For the mood analyser, Fisher Face Algorithm is used. Image recognition using Fisher face method is based on reduction of face based dimeensions and it includes Principal Component Analysis(PCA) , Fisher's Linear Discriminant (FDL) , Linear Discriminant Analysis(LDA).

6.Finally, the result will be displayed in the form of graph and efficiency in the percentage. For the graph , CanvasJS is used. It is JavaScript Charting Library.

Expected Result:

The output includes a graph, specifically a bar graph which summarizes users results in a format easy to read and understand as shown below.



FRAME WISE ANALYSIS

image90.jpg
image300.jpg
image210.jpg
image150.jpg
image180.jpg
image240.jpg
image120.jpg
image60.jpg

Applications

1]Can be used by government agency.
2]Can be used for an interview purpose.
3]Can be used by any person to prepare himself for a speech or hosting functions.
4]Can be used by college/school students for study purposes.

Future Work

Future direction is to detect subtle changes or micro expressions , as in the case of medical evaluation of depression , to further improve the classification accuracy. In future direction we see this project to be using AI for checking answer accuracy. And even direct answer comparison by parsing data available on the web.

Conclusion

This paper explains methodology that helps a user to understand its confidence and efficiency over its video by analysing it for a particular paragraph of text which gives result in graphical format representing its accuracy percentage.

References

1. “Sentence embeddings and their meanings” proposed by Nuzhat Robert.
2. “A review on speech to text conversion methods” by Prachi Khilari.
3. “Local Visual Microphones: Improved Sound Extraction from Silent Video” by Mohammad Amin
.4.”Face Recognition Using Fisherface Method” by Mustamin Anggo.
5. “Extraction of High-Resolution Frames from Video Sequences” proposed by Richard and Robert L.