International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue III, March 2018- Available at www.ijraset.com

Video based Subtitle Generation

Aditi Sankhe[1], Vishakha Patil[2], Subodh Patel[3], Siddhesh Bhagat[4]
[1, 2, 3,] *Student,* [4]*Assistant Professor Computer Engineering Department, ST John College of Engineering and Management, Palghar, Maharahstra, India*

Abstract: ***Now-a-day's video has become one of the most popular multimedia artifacts used on PCs and internet. Sound plays a vital role within a video, for example movies or the video lectures or any other multimedia data relevant to the user. Thus, Subtitles hold an important place for people with several physical problems such as auditory problems as well as for people with gaps in spoken language. Therefore most natural way lies in the use of subtitles which is done by downloading subtitles of any video from the internet is a monotonous process. This proposed system resolves the above issue by generating subtitles automatically through software and without the use of internet with the help of three distinct modules namely Audio Extraction, Speech Recognition and Synchronization. Consequently, operations are performed and subtitles generation is achieved.***
Keywords: Video, Subtitle, Audio Extraction, Speech Recognition, Synchronization

I. INTRODUCTION

In today's world Subtitles are very important to understand the content spoken by the individual in video as video plays a vital role to help people understand and comprehend information present in it. Hence, here it becomes important to play videos available to the people having auditory problems and even more for the people to remove the gaps of their native language. Thus, one of the method available for generating subtitle is manually writing the subtitles which cost us much time as well as lead to many errors and another which is simple to implement i.e downloading the subtitle file using internet. Therefore, this leads us to a valid subject of research in field of automatic subtitle generation. Thus paper provides the user a major benefit of not downloading the subtitles through internet instead generating them automatically as well as it will provide missing information for individual, who have difficulty in processing speech and auditory components of the visual media. It will establish a systematic link between the written word and the spoken word and can also be beneficial for those who learn English as second language.

II. LITERATURE REVIEW

Various research study have been done to accomplish each module of project . This compression technology is used by audio coding in sub band compression technique. The band is divided into 32 sub bands and passed through the Fast Fourier Transform operations to do the signal frequency analysis. Then the CRC, the standard MPEG stream can be obtained. At the decoder, it is only to decode the frame and the sub-band sample value. Then it does the reversion of the frequency mapping. Finally output the standard MPEG code stream.

Speech Recognition technology is used to transfer the voice signal to an associated text. This system is essentially a kind of pattern recognition system, including three basic units such as feature extraction, pattern matching and reference model library. The input speech signal characteristics are compared with the voice template stored in computer according to the speech recognition model, finding out a series of optimal template matching with input phonetic by a certain search and matching strategies. Then, computer provides recognition results by checking the table according to the template definition.

Speech Recognition is the automated conversion of speech into written text. A more speaker independent SR system while maintaining speed and accuracy of transcription. This requires the construction of an SR dictionary that takes into account the existence of multiple pronunciations for the same words. By finding the optimal number of pronunciation per word, the percentage of words correctly identified by SR system increased from 78% to 85.7%. This brings the technology 35% closer to the goal of complete recognition and the use of speech as the primary method of human-computer interaction.

III. PROPOSED SYSTEM

The proposed system starts by taking its input from the user. The input provided is the video file. This file is then passed through the decoder where the file is split into audio and video file. The audio obtain is further passed through the audio synthesizer where the audio signal is converted to the text format and then synchronized with audio in time domain. The subtitled which is generated is then passed to the miner where all three formats are clubbed together i.e. audio, video and subtitle file.

©IJRASET (UGC Approved Journal): All Rights are Reserved

The proposed system to be implemented is described briefly and can be depicted in the above diagram. It illustrates a sketch of the outline architecture. This diagram serves the general purpose to explain the broad structure and the working of the proposed system. The system consist of the following modules integrated within the framework. Command interface and UI manager : This module serves has an interface to the user. This module is basically used to accept the query from the user and display results of the same.

1) *Decoder :* In this module FFMPEG package is used for splitting the audio as well as video file from the input and also used for merging the same files so that we can get the well defined output. This package is not only used for seperating the audio from video but also converting this audio file into .wav format
2) *Audio Synthesizer :*This is main module, as all operations on audio are performed here. The .wav file is provided as input to this module . Here the audio Preprocessing block comprises of noise removal technique which will compress the .wav file. The normalization process also takes place here in which the signal is then normalized so that it can be used further. The normalized signal is then further passed through the recognition module where the audio is converted to the text format on the bases of training set provided to it. This will generate a .txt file. The. .txt file generated is uneven and non meaningful. Thus to give proper sentence formation to this file we use NLP. Hence, the output generated is give in .srt format or file. This .srt file which is generated is now synchronized with the audio using the time domain.
3) *Miner:* This module is used to club the raw video file with the synchronized audio file having subtitles.

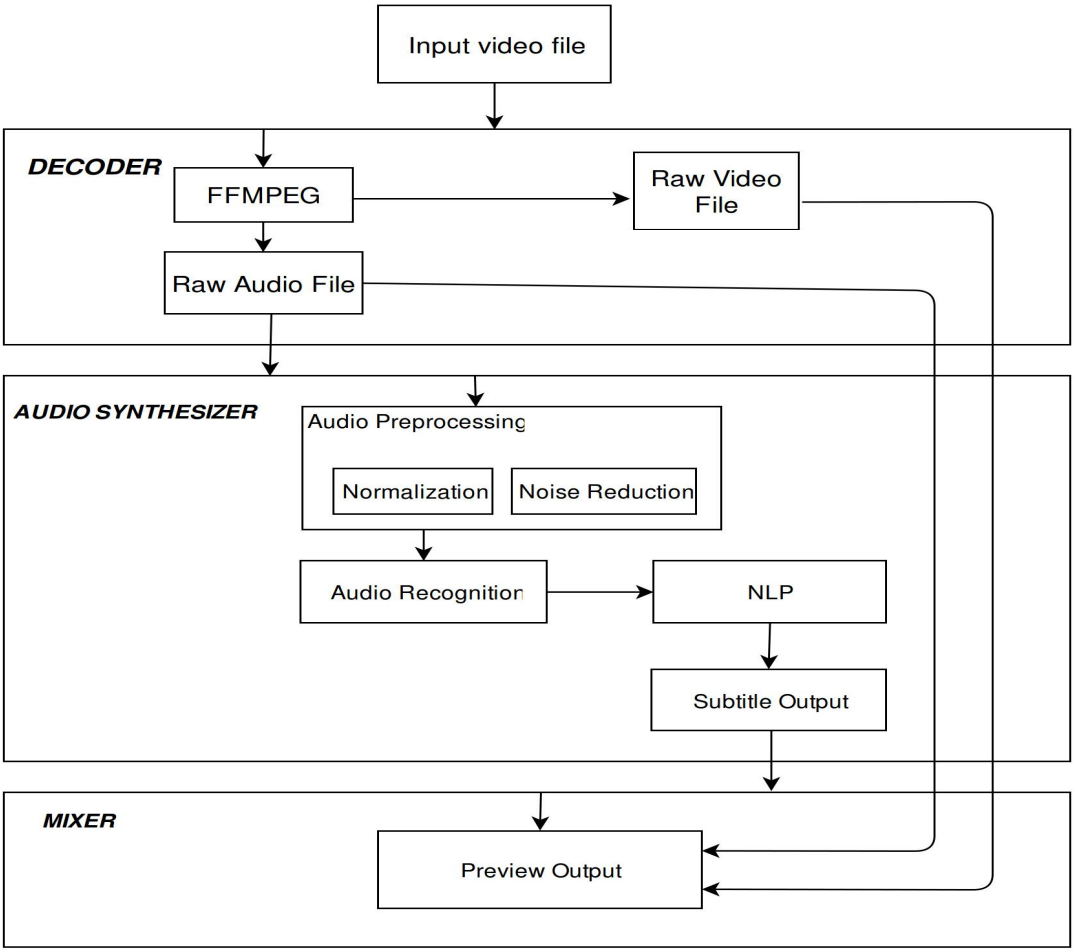

FIGURE 1: ARCHITECTURE DIAGRAM

IV. DESIGN

The unified modelling language is a standard visual modeling language intended to be used for moedeling business and similar processes, analysis, design and implementation of software based system.UML is a common language for business analyst, developers used to describe, specify, design, structure and artifacts of software system.

©IJRASET (UGC Approved Journal): All Rights are Reserved

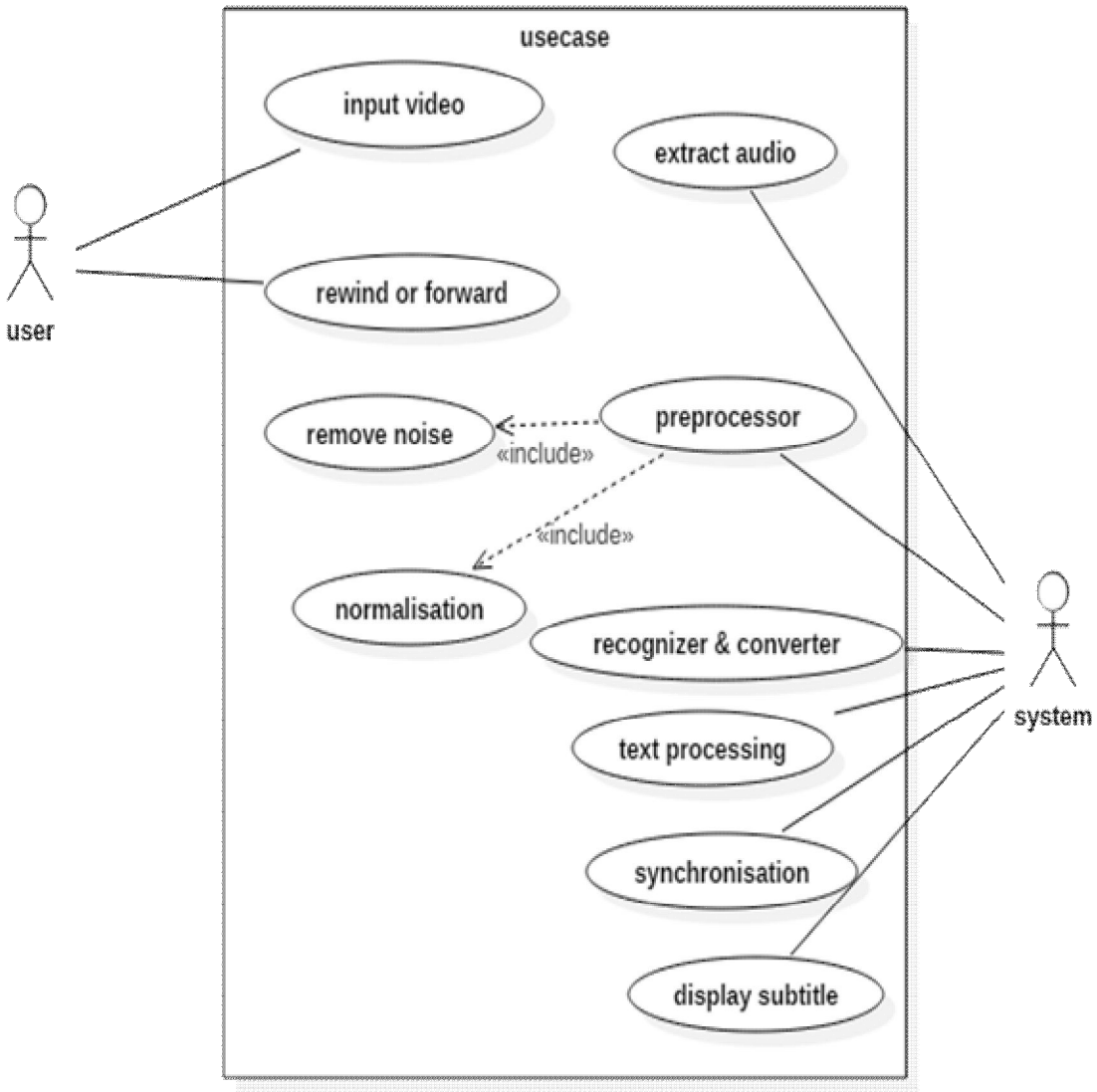

FIGURE 2: USECASE DIAGRAM

For our system we have two use-cases operator, which provides training and the user which will use the system. Operator mainly comprises of activities like audio extraction, speech recognition and NLP. Audio extraction is used to seperate the video and audio file, where as speech recognition is used for pre-processing and NLP for text generation which is sync with the audio and then displaying to the user.

V. CONCLUSIONS

This paper, we propose a video based subtitle generation. Subtitle are generated using three modules namely, Audio extraction, Speech recognition, and Synchronization. Our project overcomes the shortcomings of the previous system. We are proposing a system where subtitles are generated automatically without the use of internet. Considerably reduces time, energy wastage. Video based automatic subtitle generation is an ongoing hot research area with continuing attributions from different domains

VI. ACKNOWLEDGMENT

We are thankful to the anonymous reviewers for their valuable comments due to which the paper was improved. We are also thankful to our project guide Prof. Siddhesh Bhagat and Prof. Pawan Gujjar, Head of the Computer Engineering Department, St. John College of Engineering and Management, Palghar for their relentless support throughout the period of the project

©IJRASET (UGC Approved Journal): All Rights are Reserved

VII. RESULT

Our result involes a user interface

REFERENCES

[1] Hongz zohu and chang hui You, "Research and design of the audio coding scheme",IEEE transactions of a consumer Electronics, international conference on multimedia technology 2011

[2] Justine Burdick, "Building a Regionally inclusive Dictionary for speech Recognition,"computer science and Linguistic, spring 2004.

[3] Youhao Yu “Research on Speech Recognition Technology and Its Application,” Electronics and Information Engineering, International Conference of Computer Science and Electronics Engineering, 2012.

[4] Sadaoki Furui, Li Deng, Mark Gales,Hermann Ney, and Keiichi Tokuda,, ” Fundamental Technologies in Modern Speech Recognition,” Signal Processing, IEEE Signal Processing Society, November 2012.

[5] Yu Li, LingHua Zhang, “Implementation and Research of Streaming Media System and AV Codec Based on the Handheld Devices,” in 12th IEEE International Conference on Communication Technology (ICCT), 2010.

[6] Jing Wang, Xuan Ji, Shenghui Zhao, Xiang Xie and Jingming Kuang, “Context-based adaptive arithmetic coding in time and frequency domain for the lossless compression of audio coding parameters at variable rate,” EURASIP Journal on Audio, Speech, and Music Processing 2013.

[7] Anand Vardhan Bhalla, Shailesh Khaparkar, “Performance Improvement of Speaker Recognition System,” International Journal of Advanced Research in Computer Science and Software EngineeringVolume 2, Issue 3, March 2012.

©IJRASET (UGC Approved Journal): All Rights are Reserved