Automatic Subtitle Generation for Videos

Akhil Kanade, Sourabh Gune, Shubham Dharamkar, Rohan Gokhale

PES's, MCOE Department Of Computer Engineering, Pune-05, sourabhgune@ymail.com, Mob.:9765353586

Abstract— The main objective of developing this system is to present an automated way to generate the subtitles for audio and video. By replacing the tedious method of the current system will save time, reduce the amount of work the administration has to do and will generate the subtitles automatically with electronic apparatus. This system will first extract the audio, then recognise the extracted audio with the available speech recognition API. Later the recognized audio is converted to the text and saved in text file having extension ".srt". Later on, this ".srt" file can be opened in a media player to view the subtitles along with video.

Keywords—Audio extraction, Speech recognition, .srt file, Time synchronization, Automatic Subtitle generation, Natural language processing.

INTRODUCTION

About this Project:
The main idea of developing this system is to present an automated way to generate subtitles using audio extraction and speech recognition techniques which would replace the present method of writing the file manually. Replacing the tiresome old method this system will save the time, reduce the amount of work the administration has to do and will minimize the human errors associated with this process.
Motivation:
Nowadays due to increase in use of syllabus related videos for teaching in classes, may that be at school or college level, some students are unable to grasp what the speaker is trying to explain in that video. If the videos are shown along with captions then it becomes easy to relate what the video or the speaker in video want to convey. Also the videos are not compulsorily provided with subtitles so manually write this is impossible task for any individual, also to search the respective time synchronized file may take time. Instead by using this software any individual can easily generate the captions and club it with video which can help students and all.

BRIEF DESCRIPTION

Need of Automatic subtitle generation:
Subtitles are very important to understand the content spoken by the individual in video. There is an alternate method available for generating subtitles i.e. manually writing the file but it costs us much time. This method waste time because the individual has to manually write the subtitle file which is a tedious task which may introduce many errors also the time synchronization must be done which every individual is unable to perform. This work describes the efficient process that automatically generates the subtitle without human intervention. This subtitle is generated by using a speech API and the file is produced which is properly time synchronized and displays accurate subtitles.
In today's world use of subtitles has become important for understanding of the video. Subtitling is essential for people who are deaf those who have reading and literacy problems and can to those who are learning to read. Subtitles provide information for individuals who have difficulty understanding speech and auditory components of the visual. This leads to a valid subject of research in field of automatic subtitles generation. Thus this report provide users a major benefit of not downloading the subtitles instead generating them automatically. Various studies have been done to accomplish this type of process. Downloading the subtitles from the internet is a monotonous processes. Consequently, to generate the subtitles thorough the software itself is a easy process.

Innovativeness and Usefulness:-
1) The major benefit is that the viewer does not need to download the subtitle from Internet if he wants to watch the video with subtitle.
2) Captions help children with word identification, meaning, acquisition and retention.
3) Captions can help children establish a systematic link between the written word and the spoken word.
4) Captioning has been related to higher comprehension skills when compared to viewers watching the same media without captions.
5) Captions provide missing information for individuals who have difficulty processing speech and auditory components of the visual media.

6) Captioning is essential for children who are deaf and hard of hearing, can be very beneficial to those learning English as a second language, can help those with reading and literacy problems and can help those who are learning to read.

This project will mainly generate subtitles with help of speech API.
Generation of subtitles include following steps:

Speech Extraction:

The audio extraction routine is expected to return a suitable audio format that can be used by the speech recognition module as pertinent material. It must handle a defined list of video and audio formats. It has to verify the file given in input so that it can evaluate the extraction feasibility.

Speech recognition:

The speech recognition routine is the key part of the system. Indeed, it affects directly performance and results evaluation. First, it must get the type of input file then, if the type is provided an appropriate processing method is chosen. Otherwise, the routine use a default configuration.

Subtitle Generation:

The subtitle generation routine aims to create and write in a file in order to add multiple chunks of text corresponding to utterances limited by gaps and their respective start and end times. Time synchronisation consideration are of main importance.

For analyzing Extraction and Recognition, we have to take following concepts into consideration as follows:

1) Recognizing different speakers:

Many instances are observed where there are multiple sources of sound which may give error in subtitles. The main challenge is to identify the speaker for whom the caption need to be generated.

2) Recognizing the silences/gaps:

Many pauses are observed when there is verbal conversation between speakers. This situation must be correctly analyzed so as to obtain correct output.

3) Generation and appending of text:

Generated sentences or words must be appended in proper sequence into the text file.

4) Time synchronize the text:

The biggest challenge is to display the captions according to the video. The text file contains the sentences corresponding to the input video which must be time synchronized according to the frames.

Major constraints:

Accuracy:

Accuracy of system should be high otherwise the generated subtitles will be incorrect for the purpose. The process of extraction and recognition must be correctly followed.

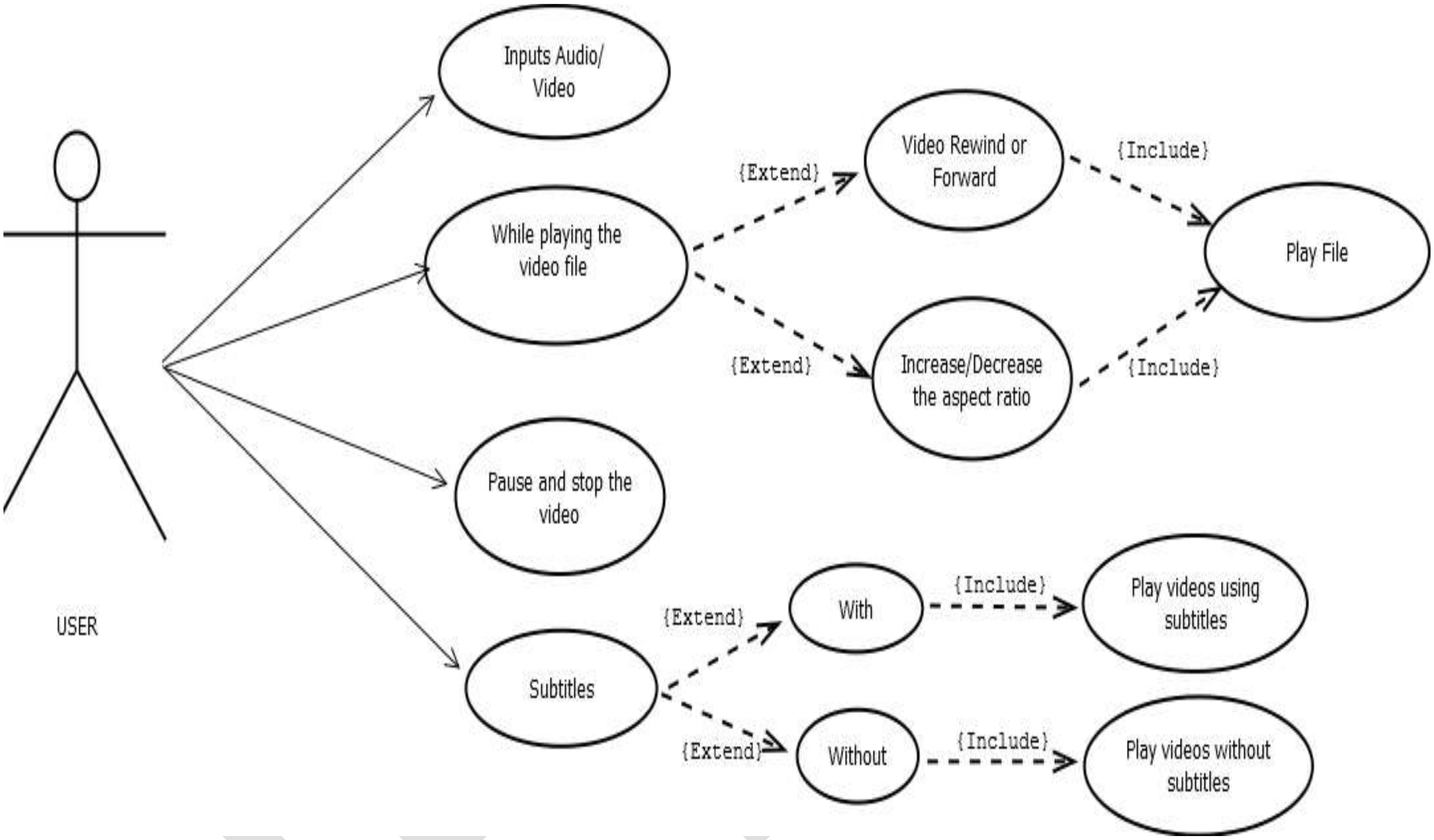
Input Video:

The quality of video which is given as input must be good with less audio distortions.

Speech:

During the process of audio extraction the audio must be properly extracted as there maybe loss of contents which may produce errors at later stages.

Use case diagram:



CONCLUSION

By using this software subtitles or subtitle file will be generated for any English videos. This software will minimizes the efforts for downloading or manually writing the subtitle file. Any one will be able to generate the subtitle file as this software is very easy to use and just needs input which can be provided by anyone. Also this software can be used with online website to provide videos along with subtitles. Many type of formats will be supported by this software.

REFERENCES:

[1] Abhinav Mathur, Tanya Saxena, Generating Subtitles Automatically using Audio Extraction and Speech Recognition, 7th International Conference on Contemporary Computing (IC3), 2015.

[2] 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.

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

[4] Jorge Martinez, Hector Perez, Enrique Escamilla, Masahisa Mabo Suzuki, Speaker recognition using Mel Frequency Cepstral Coefficients (MFCC) and Vector Quantization (VQ) Techniques, 22nd International Conference on Electrical Communications and Computers (CONIELECOMP), 2012

[5] Anand Vardhan Bhalla, Shailesh Khaparkar, Performance Improvement of Speaker Recognition System,International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 3, March 2012

[6] Ibrahim Patel Dr. Y. Srinivas Rao, Speech Recognition Using HMM with MFCC- An Analysis using Frequency Spectral Decomposition Technique Signal and Image Processing: An International Journal(SIPIJ), Vol.1, No.2, December 2010.

[7] B. H. Juang; L. R. Rabiner, “Hidden Markov Models for Speech Recognition” Journal of Technometrics, Vol.33, No. 3. Aug., 1991.

[8] Hong Zhou and Changhui Yu , “Research and design of the audio coding scheme ,” IEEE Transactions on Consumer Electronics, International Conference on Multimedia Technology(ICMT) 2011.

[9] Seymour Shlien,”Guide to MPEG-1 Audio Standard”, Broadcast Technology, IEEE Transactions on Broadcasting, December 1994.

[10] Justin Burdick, “Building a Regionally Inclusive Dictionary for Speech Recognition”, Computer Science and Linguistics, Spring 2004.

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

[12] Ibrahim Patel1 Dr. Y. Srinivas Rao, “Speech Recognition Using HMM with MFCC- An Analysis using Frequency Spectral Decomposition Technique”, Signal & Image Processing: An International Journal(SIPIJ), Vol.1, No.2, December 2010.

[13] Stephen J. Wright, Dimitri Kanevsky, LiDeng, Xiaodong He, Georg Heigold, , and Haizhou Li, “Optimization Algorithms and Applications for Speech and Language Processing,” IEEE Transactions on Audio, Speech and Language Processing, Volume 21, Issue 11, November 2013.

[14] 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.