Scientific Journal Impact Factor (SJIF): 3.518

IJMTER

International Journal of Modern Trends in Engineering and Research

ISSN (ONLINE) : 2349 - 9745

ISSN (PRINT) : 2393 - 8161

Generating Subtitles Automatically For Sound in Videos

Prof. S. A. Aher [1] , Hajari Ashwini M [2] , Hase Megha S[3], Jadhav Snehal B[4] ,Pawar Snehal S.[5]

[1] *Asst. Prof., I.T. Dept., S.V.I.T. Nashik, India*

[2,3,4,5] *B.E. Students, I.T. Dept. , S.V.I.T. Nashik, India*

Abstract— Now-a-days, video plays an important role to understand the information to people more clearly and to gain deep knowledge and some of its examples are the song, movies or the video lectures. Certain Individuals cannot understand the meaning of such videos because there is not any text transcription available. Hence, it is important to make videos available for people having auditory problem and people who don't have the deep knowledge of English language and also to remove the gaps of their native language, which can be done by using Subtitles of the videos. Earlier, people need to download subtitles of any video from internet which is very problematic. Hence, generating subtitles automatically without any use of internet is very interesting subject to research. This paper include the same concept with three important steps namely Audio Extraction, Speech Recognition and finally, Subtitle Generation.

Keywords— video; subtitles; Audio Extraction; Speech Recognition; Subtitle Generation.

I.INTRODUCTION

Today, the use of subtitles is very important for understanding videos. It is very helpful for people who are deaf, who have reading and literacy problems also to those who are learning to read[4]. By providing subtitles, helps the people to understand the speech and auditory components of the visual, which leads to a valid subject of research in the Automatic Subtitle Generation field. However, manually creation of Subtitles is a very long and boring activity which requires an extensive presence of user is essential. This paper includes the user a major benefit of not to download the subtitles from the internet instead it provide a deep information about generating Subtitles Automatically.

At present, in VLC (Video LAN Client) media player the subtitles must have to be inserted first to media player then it synchronized with the song. The inserted file contains .srt file which contains the time intervals of text spoken in the given input file instead of .txt file which do not contain the time intervals of the spoken words. While YouTube accepts both .txt file and .srt file for synchronization [1].

Various studies have been done to accomplish each module of the project but by using Speech Recognition Generation of Subtitles have not been developed which is our main motive. HMM (Hidden Markov Model) is used for Speech Recognition which calculates probability of occurrence of words by using Language model and the Acoustic [3]. This paper describes the techniques to Generate Subtitles Automatically with detailed description of three modules namely, Audio Extraction, converts an input file supported by MPEG standards to .wav format, Speech Recognition of extracted .wav file is implemented and finally, Subtitle Generation in which a .txt/.srt file is generated which is synchronized with the given input file

II.LITERATURE SURVEY

In 2007, Automatic Video Classification paper was published by Darin Brezeale and Diane J.Cook, Senior member IEEE includes, some features from three modalities such as Text, Audio and Visual which is combination of features and classification. This paper gives brief explanation about general features and summarize the research in this area.[5]

The Structured Discriminative Models for Speech Recognition by Mark Gales and Eric Fosler-Lussier Senior Members, IEEE includes detailed description of Structured Discriminative Models for Recognition of Speech. The model handle Variable- Length Observations and Word (sub-sentence) sequences which is generally used for task processing in natural language which not required account for segmentation and it is introduced for ASR. Model parameters are tied together and robustly estimated by using Segmentation of the sentence.[6]

Hidden Markov Model for Speech Recognition by D.B.Paul, the model gives stochastic from which is flexible but rigorous to build the system. It also provide the training algorithm for estimating the parameters of the model which is computationally efficient. Some improvements are needed such as speaker-dependent and speaker independent in continuous speech recognition. During this research, the model is not give the good performance on current 1000 words vocabulary with the available technologies that time.[7]

In previous system manual method is used to create subtitle by using voice recognition software, which dictate subtitle to a given text file without considering the time specification gives number of subtitle and subtitle. Afterwards text file is being save and open file in text editor. Save the text file with ".SRT" file extension by formatting each subtitle in four fine format. To drag and drop subtitle into timeline of video by using a caption and subtitling tool along with Graphic user interface, such as InqScribe or Viddler Aegisub. For every subtitle in time after applying start and stop controls, the location of each entry and proper format of time specification are recorded to the SRT file. ".SRT" file stands for "SubRip Text file", contains simple formatted text data which eas first written in france. The start and end time are written in "hours:minutes:seconds:miliseconds" format.

III.PROPOSED SYSTEM

A. Proposed System Architecture:

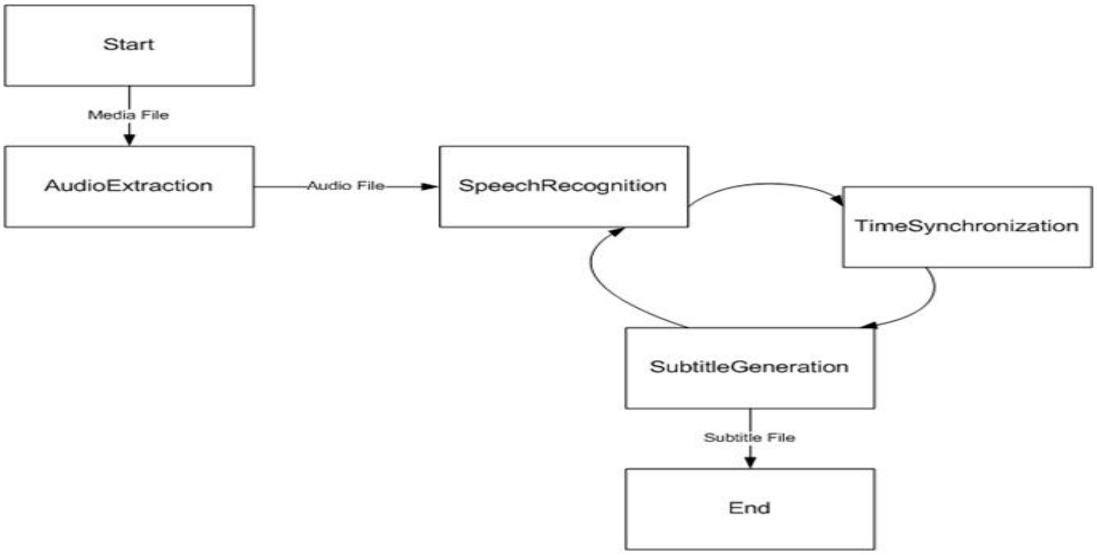

Figure 1: Proposed System Architecture.

Diagram given above shows the system architecture of Generating Subtitle Automatically for Sound in Videos in which a Media file either Audio or Video is given as input to system. Track of Audio is extracted and chunk is read line by line until the end of track is reached. For successful processing three tasks are going on as Speech Recognition, Time Synchronization and Subtitle Generation which finally generate a file as output.

B. Framework

Three distinct modules have been defined as, Audio Extraction, Speech Recognition and Subtitle Generation which take a video file as a input and generate subtitle file as output.

1) Audio Extraction:

Audio extraction model is expected to return a suitable audio format that used by the speech recognition module as pertinent material. List of video and audio formats must be handle by it and verify file given as input which can evaluate the extraction feasibility.

2) Speech Recognition:-

It is the key part of the system which affects directly to performance and results evaluation. After identification of input file then, if the type is provided, an appropriate processing method is chosen otherwise, the routine uses a default configuration.

1) File type is specified like WMV, AVI, MP4.
2) After type specification, appropriate method is chosen.
3) If that is not specified default configuration is accessed.
4) Modern Speech recognition system is based on Hidden Markov Model(HMMs) Algorithm.

3) Subtitle Generation:-

Multiple chunks of text corresponding to utterances limited by silences and their respective start and end times create and write in a file.

1) We get a list of words and their respective speech time from speech recognition module.
2) After that to produce a SRT subtitle file.
3) Module checks for list of words and use silence utterances as delamination for consecutive sentences.

C. Proposed System Algorithm

a) Audio Extraction:

1) **Input:-** Takes a Video File
2) **Procedure:-** Checks the file content, creates a list of separated tracks composing the initial media file. Here audio will be extracted from media file i.e. from video.
3) **Output:-** Get an Audio File

b) Speech Recognition:-

1) **Input:-** An Audio File
2) **Procedure:-** Though a line tool audio file which we have created it will be a output to google API which is used to generate text. This required Internet connection.
3) **Output:-** The Recognized speech(list of words).

c) Subtitle Generation:-

1) **Input:-** Give the List of words and their respective speech time.
2) **Procedure:-** The text which will be generate in chrome browser will be saved in a file “.text”.
3) **Output:-** Generate the SRT subtitle file.

d) Time Synchronization:-

1) **Input:-** The List of Words.
2) **Procedure:-** It makes synchronization between spoken audio and generated subtitle file.
3) **Output:-** The List of words and their respective speech time.

IV. CONCLUSION

In this paper, we proposed a way to generate Subtitle automatically which includes three modules as Audio Extraction audio is extracted into .wav file format from given input video. The average size of input file taken is reduced up to 10MB to 12MB after Extraction. The average bitrate input before extraction that reduces after extraction. In Speech Recognition, the extracted .wav audio file is decoded. Text of song is recognized using the Acoustic model which shows the occurrences of words and .srt file is generated which contains the text of input file.

In the World of Internet, it is essential to give each individual the right t understand any media content. The Internet has known as multiplication based on videos used by many websites which are rarely available.

This paper contains the concept of Automatic Subtitle Generation is explained in details as it includes the completion of 70-75% of the project with Audio Extraction and Speech Recognition. We are now working on the last and final module of the project which is Automatic Subtitle Generation which is the further work that could be realized to enhance the concept.

V. ACKNOWLEDGEMENT

A dissertation of this magnitude has been a journey with various ups and downs. Whenever We are standing on most difficult step of the dream of our life, We often remember the great almighty god for his blessings kind help and he always helps us in tracking off the problems by some means on our lifetime. We feel great pleasure to represent this project by using Subtitle Generation Automatically. We would like to convey sincere gratitude to our project guide and head of information technology engineering department Prof. S.A Aher for her valuable guidance and support and who guided us and provided with her useful and valuable suggestions and without her kind co-operation it would have been extremely difficult for us to complete this dissertion. We would like to convey sincere gratitude to B.E. Co-cordinator Prof. R.S. Bhalerao for his valuable guidance and support. We would like to express our appreciation and thanks to Prof .Dr. G. B. Shinde, Principal, S.V.I.T. Nashik.

Finally we are very grateful to inspiring parents who loving and caring support contributes a major share in completion of our task

REFERENCES

[1] International Conference on Computational Intelligence and Communication Technology 2015 IEEE Transaction for Generating Subtitles Automatically using Audio Extraction and Speech Recognition

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

[3] 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.

[4] Youhao Yu "Research on Speech Recognition Technology and its Application," Electronics and Information Engineering, International Conference on Computer Science and Electronics Engineering, 2012.

[5] Automatic Video Classification: A survey of the Literature[Senior member, IEEE 2007].

[6] Structured Discriminative Models for speech Recognition[Mark Gales, fellow,IEEE,Shinji watanabe,senior member, IEEE, eris fosler Lussier Senior Member, IEEE].

[7] Speech Recognition using Hidden Markov Models[D.B.Paul 1990].

[8] G. Atenises, A.D. Santis, A. L. Ferrara, and B. Masucci, provably-Secure Time-Bound Hierarchichal Key assignment Schemes, J. Cryptology, vol.25, no.2,pp.243270,2012