Multimedia Summarization using Audio Descriptors

[1]Prof. Mrs. A. R. Kulkarni , [2]Prof. Dr. Mrs. S. S. Apte , [3]Mr. H.M. Tamboli

[1]Assistant Professor in Department of Computer Science & Engineering
Walchand Institute Of technology, Solapur, India

[2]Professor in Department of Computer Science & Engineering
Walchand Institute Of technology, Solapur, India

[3]M.E.C.S.E. Department of Computer Science and Engg.
Walchand Institute of Technology, Solapur, India.

ABSTRACT

Video summarization is defined as creating a shorter video clip which includes only the important scenes in the original video streams. It has a wider research area since it is more complex to implement. Even by implementing all the serious algorithms, the success levels can vary on the complexity of the video.Currently video summarizing is human-interacted, the person who is dealing with the raw video should pick highlights of each video and join them together to generate a series of highlights. This project tries to address this problem domain by summarizing a video.In this project a novel approach is proposed to multimedia summarization using the audio descriptor. This uses the audio descriptor from a given video, a system will then generate a paragraph of textual information of the video. By summarizing the textual information & accessing the video content relevant to the summarized output, get the summarized video of the desired length. This summarization is simple and flexible, which allows rapid generation of a summary of any desired length.

Keywords: multimedia summarization, HITS algorithm ,audio descriptor.

1. INTRODUCTION

The rapid development of digital video capture and editing technology led to increase in video data, creating the need for effective techniques for video retrieval and analysis. Advances in digital content distribution and digital video recorders, has caused digital content recording easy. However, the user may not have enough time to watch the entire video. In such cases, the user may just want to view the abstract of the video instead of watching the whole video which provides more information about the occurrence of various incidents in the video.As the name implies, video summarization is a mechanism for generating a short summary of a video, which can either be a sequence of stationary images (key frames) or moving images (video skims) [2]. Video can be summarized by two different ways which are as follows.

1) Key Frame Based Video Summarization

These are also called representative frames, R-frames, still-image abstracts or static storyboard, and a set consists of a collection of salient images extracted from the underlying video source [2]. Following are some of the challenges that should be taken care while implementing Key frame based algorithm

1. Redundancy: frames with minor difference are selected as key frame.

2. When there are various changes in content it is difficult to make clustering.

In this paper, we will carry video summarisation work into three stages: First we will take the audio descriptors from the video file. In the second stage, we will extract the text data from the audio descriptors file. Graph-based ranking algorithms, such as Kleinberg's HITS algorithm or Google's PageRank have been traditionally and successfully used in citation analysis, social networks, and the analysis of the link-structure of the World Wide Web. In this paper, we investigate a range of graph based ranking algorithms, and evaluate their application to automatic unsupervised sentence extraction in the context of a text summarization task. We will use the Graph based algorithm to create the summary of the obtained text information of the video. In the third stage, we will extract the video files related to the resulted summary. Finally, a summary would describe the videos, and explain what the information does the video contains.

When you submit your paper print it in one-column format, including figures and tables. In addition, designate one author as the "corresponding author". This is the author to whom proofs of the paper will be sent. Proofs are sent to the corresponding author only.

International Journal of Application or Innovation in Engineering & Management (IJAIEM)
Web Site: www.ijaiem.org Email: editor@ijaiem.org
Volume 4, Issue 6, June 2015   **ISSN 2319 - 4847**

2. LITERATURE REVIEW

2.1 Audio Summarization

Automatic audio summarization is an on going research pursuit, which relies either on algorithms to identify and remove redundancy, for example in music or noise, or first turn speech into text, and then employ text summarization methods. The peculiarities and potential ambiguities of decoded audio such as high recognition error rates, lack of syntactic boundaries, etc., need to be addressed specifically for extracting summary information from audio for content-based browsing and skimming. Generally, speech summarization can be performed by simply extracting salient words or sentences from the original data, or by synthesizing new representations from the original data. The second case is of course more difficult, harder to evaluate, but also potentially more useful, because the information representation cannot only be compact, but also targeted, clean, and easy to understand. It is our goal for multimedia summarization to achieve similar progress with respect to video retrieval.

2.2 Text Summarization

Text summarization is the process of automatically creating a compressed version of a given document preserving its information content. Text summarization can be categorized into two approaches: extractive and abstractive. Extractive summarization methods simplify the problem of summarization into the problem of selecting a representative subset of the sentences in the original documents. Algorithms for extractive summarization are typically

based on techniques for sentence extraction, and attempt to identify the set of sentences that are most important for the overall understanding of a given document.Abstractive summarization may compose novel sentences, unseen in the original sources. However, abstractive approaches require deep NLP such as semantic representation, inference and natural language generation, which have yet to reach a mature stage now a days. Extractive summarization systems are commonly used in automatic summarization to produce extractive sum-maries. Systems for extractive summarization are typically based on technique for sentence extraction, and attempt to identify the set of sentences that are most important for the overall understanding of a given document. Most commonly, such systems use some kind of similarity or centrality metric to identify the set of sentences to include in the summary

2.3 Video Summarization

A video summarization is a summary which represents abstract view of original video sequence and can be used as video browsing and retrieval systems. It can be a highlight of original sequence which is the concatenation of a user defined number of selected video segments or can be a collection of key frames. Different methods can be used to select key frames. key frames based video summarization works on frames so first step is to extract frames from original video frame sequence. In step two extracted video frames are cluster that have redundant content obviating the need for shot detection. Selection of key frames is proceeding in step three. The entire procedure is shown in fig.
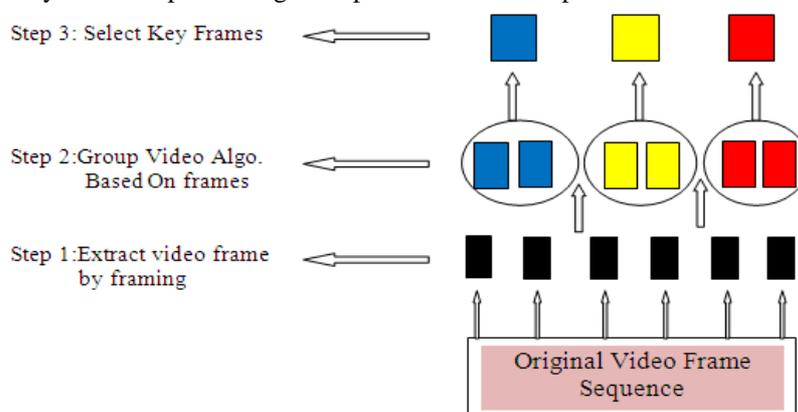
Fig.2: Key frames based video summarization

Key frames based video summarization can be classified in three different ways. These are as follows.

1) Classification based on sampling

This method [11] chooses key frames uniformly or randomly under-sampling, without considering the video content. The summary produced by these methods does not represent all the video parts and may cause some redundancy of key frames with similar contents.

2) Classification based on scene segmentation

This method [11] extracts key frames using scenes detection, the scene includes all parts with a semantic link in the video or in the same space or in the same time. The disadvantage of these techniques is producing a summary, which does not take into account the temporal position of frames.

3) Classification based on shot segmentation

This method [11] extracts adapted key frames to video content. They extract the first image as shot key frames or the first and the last frames of the shot. These methods are effective for stationary shot and small content variation, but they don't provide an adequate representation of shot with strong movements.

3. METHODOLOGY
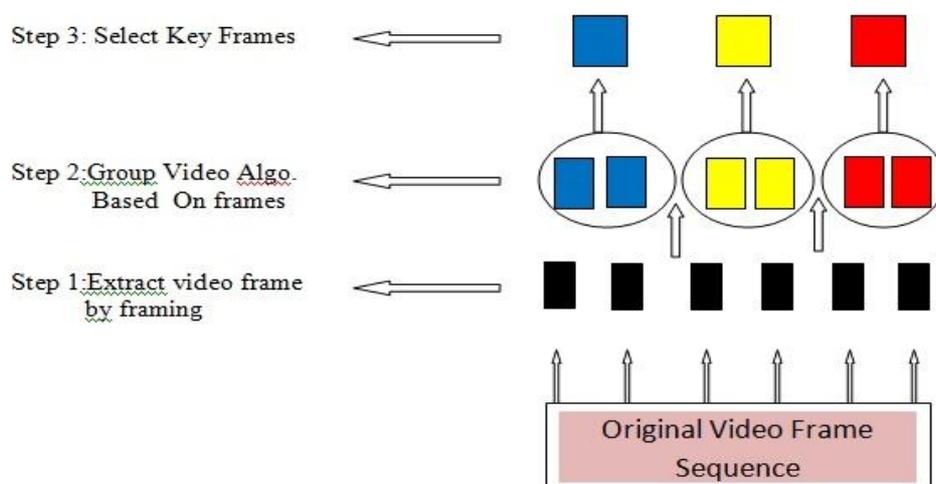
Figure 1: Keyframe based vide summarization.

Above we present the flow of our proposed work, we will take the multimedia requirement of the user which can be used as input to the model. Then we will access the required data from the server. The obtained data consists the video data, audio data and audio descriptors. Our interested area is the audio descriptor files. Audio descriptor contains the audio text data, timestamp and other optional parameters. We will extract the text data from the audio descriptor. According to the information obtained by the audio descriptor, we will collect all the text data from audio descriptors. After obtaining whole data, we will summarize the data. For summarization purpose, we will use Extractive method. From the resulted text summary, we will access the video files relevant to the obtained summary. So, this summarized multimedia information will explain the information of the video shortly. So, user not needed to watch the video completely. From the summary, he can understand what the video is contains.

4. CONCLUSION

We proposed to show the summarization of video using audio descriptor. In this application we will provide such a technique which provide summarized data for given video. We will differentiate our work from prior systems in that we are investigating static summaries with a multimedia component, i.e. ones that can be viewed all at once, rather than playing a video from start to the end. We will also show that a summarized data can be generated in such a way that a user can get the idea about the video without watching the full video.

REFERENCES

[1]. Rada Mihalcea: Graph-based Ranking Algorithms for Sentence Extraction, Applied to Text Summarization
[2]. S. Brin and L. Page. 1998. The anatomy of a large-scale hypertextual Web search engine. Computer Networks and ISDN Systems, 30(1–7). DUC. 2002. Document understanding conference 2002. http://wwwnlpir.nist.gov/projects/duc/.
[3]. P.J. Herings, G. van der Laan, and D. Talman. 2001. Measuring the power of nodes in digraphs. Technical report, Tinbergen Institute.
[4]. J.M. Kleinberg. 1999. Authoritative sources in a hyperlinked environment. Journal of the ACM, 46(5):604–632.
[5]. C.Y. Lin and E.H. Hovy. 2003a. Automatic evaluation of summaries using n-gram co-occurrence statistics. In Proceedings of Human Language Technology Conference (HLT-NAACL 2003), Edmonton, Canada, May.
[6]. C.Y. Lin and E.H. Hovy. 2003b. The potential and limitations of sentence extraction for summarization. In Proceedings of the HLT/NAACL Workshop on Automatic Summarization, Edmonton, Canada, May.
[7]. R. Mihalcea and P. Tarau. 2004. TextRank – bringing order into texts.
[8]. Shaoshuai Lei, Gang Xie, and Gaowei Yan: A Novel Key-Frame Extraction Approach for Both Video Summary and Video Index1
[9]. J. Son, H. Lee, and H. Oh, "PVR: a novel PVR scheme for content protection," IEEE Transactions on Consumer Electronics, vol. 57, no. 1, pp. 173–177, 2011.
[10].B.T.Truong and S. Venkatesh, "Video abstraction: a systematic review and classification," ACM Transactions on Multimedia Computing, Communications and Applications, vol. 3, no. 1,