

A Unified Framework for Summarization of Nollywood Movie Sequences and Audio Sound Data

Ogar Tumenayu Ofut
Department of Computer
Science,
Cross River University of
Technology Calabar,
Nigeria

Francis Bakpo
Sunday
Department of Computer
Science
University of Nigeria
Nsukka, Nigeria

Hyacinth Agozie
Eneh
Department of Computer
Science
University of Nigeria
Nsukka, Nigeria

Umoh Enoima Essien
Department of Computer
Science,
Cross River University of
Technology Calabar,
Nigeria

ABSTRACT

In recent times, the robustness of movie sources for information and consumptions has attained an unprecedented height. If a successful extraction framework of valuable and interesting information from a movie without the viewer going through the entire movie to understand the storyline is obtainable that is referred to as automatic movie summarization. This act of automatic summarization can foster the management of the growing volume of movie data. This paper shall border on the design of a unified framework for summarizing both movie sequences and its corresponding audio sound data for an automatic summarization task of Nollywood movies. The framework is a Recurrent Neural Network (RNN) model with two layers that handles the selection of key movie sequences, while the audio sound processing and concatenating selected keyframe is handled by Fast Forward Moving Picture Expert Group (FFMPEG). The proposed framework produced good results on the two benchmark datasets, which shows that the experimentation of the new model is properly executed.

Keywords

Summarization, Movie, Fast Forward Moving Picture Expert Group, Neural Network, Sequences, Sound Data, Audio, Shorts, Frames.

1. INTRODUCTION

The size of the Nigerian movie industry known as Nollywood has hovered around N9bn for the past several years placing it to be the second highest revenue earner in present day Nigeria. The twenty-first century Nigerian movie industry (Nollywood) produces about 2,000 movies a year, which has ended her the third position on the global movie ranking [15]. The Nollywood industry today is the biggest film industry in Africa and one of the leading industries in the world with respect to the number of films produced per year [1]. Thus, stakeholders in the movie industry are overwhelmed with an enormous and increasing amount of movie information, which often makes it very difficult to management, search, and retrieval of specific content.

It would be practically impossible to watch all home movies produced by Nollywood thoroughly and arrange them according to their respective categories and subject matter which is extremely important when searching for a specific movie. Currently, this categorization in Nollywood movie industry is dependent on movie preview or trailers.

Movie previews are intended to get the target audience excited about seeing a movie at the cinema. Movie trailers are growing more important in the marketing of films because in

the past they were typically confined to theaters and screened during the previews for upcoming attractions [2]. However, previews are also a form of condensed storytelling. Since the aim of the preview is to attract an audience to the movie, these excerpts are usually extracted from the most interesting part of the movie. Movie previews are one form of media that can easily attract young people's desire to see a given film. Nollywood movie preview are broadcasted on TV, on the Internet, and in theaters, and they present a fascinating nature of movies through their pictures.

Currently, the existing process of making movie preview in Nollywood go through the following steps; first, a film distribution company provides an advertising concept for the movie. Next, a movie preview production company is asked to make a preview. The producer in charge of the preview then films a demo version manually, which is sent to the distribution company for screening. The above steps are repeated until the distribution company gives its final approval. Then, the sound and picture are properly edited, and the movie preview is complete.

The existing process (movie preview) for boosting users/viewer's interest by Nollywood movie industry is highly unreliable in application and hence a better way is required to create a summarized representation of the movie that is easily comprehensible in a short amount of time.

Nollywood directors, producers, distributors, marketers and all other stakeholders should rise above the mentioned challenges in order avoid the implications of being flooded by a digital revolution. The movie producer may not have the time to go through the long movie content at all time to satisfy the requirement for creating previews. The major implication of the current system to stakeholders are poor time efficiency, little or no transparency and lack of clearness. Modern digital technologies and skills upgrading are required for a more robust marketing communications in the twenty first century in order to compete with the global society.

In this paper, the focus is on building an intelligent framework with the ability to understand how audio sound data and movie sequences can be extracted simultaneously from the full movie to an interesting summary. The proposed framework is targeted at producing a Movie sequences summarization system that creates a short summary of the content of a full movie document by selecting the most informative or interesting portion for potential users/viewers.

2. RELATED WORKS

Since the use of movies in the web has drastically increased over the years, a good number of studies and the application

based on movie summarization techniques have been proposed and it has been considered from different perspectives. The motive behind movie summarization is to obtain a diverse subset of movie frames or key shots that are not only related to the given query but also contains the original information of a movie. It can be seen in two perspectives: (1) Unsupervised movie summarization (2) Supervised movie summarization.

Unsupervised movie summarization

Unsupervised movie summarization approaches the following author took advantage of in the corresponding selection criteria for summaries include; content frequency [5,6], coverage [7,8] relevance [9,10,11], and user’s attention [12], etc. following these different criteria, several approaches have been developed. Among them, clustering-based methods are the most obtainable ones [6,7]. It clusters the visually similar frames or shots into groups, in which the group centers are considered as the representative elements of the video and therefore selected as the keyframes or keyshots. Dictionary learning techniques is also another popular technique used in unsupervised video summarization [7, 8]. Unsupervised approaches are a very popular approach in video summarization both in time past and present. They are particular at making the summarization meet the required needs, they are proven to be consistent, representative, and informative. For this task, a novel machine learning model for unsupervised Nollywood video summarization was adopted.

Supervised movie summarization

Supervised methods of movie summarization use train classifiers to learn the importance of a frame for a summary. The process always commences with the segmentation of movie frames in a uniform manner into an equal size of sorts,

as done by Gygli et al. [12], or using algorithms like kernel temporal segmentation (KTS) by Potapov et al. [11]. Gygli et al. [12] computed an interestingness score for each segment using a weighted sum of features by combining low-level spatio-temporal saliency or high-level motion information, while [3] measured frame-level importance using learned factorization. Another approach is suggested by [11] to train SVMs to classify frames in segments obtained through KTS.

Recurrent Neural Networks (RNNs)

A recurrent neural network (RNN) is a type of artificial neural network which uses sequential data or time series data. RNN is a neural network is known for analyzing streams of data by means of hidden units. Recurrent neural networks (RNN) take input at multiple time steps and analyses the sequential patterns from the previous data unlike traditional deep neural networks which consider only a single input [13]. Since RNNs deal with sequential data, they are well suited for the health informatics domain where enormous amounts of sequential data are available to process [14].

3. SYSTEM ARCHITECTURE

The architecture of the proposed RNN model for Nollywood movie sequences summarization is shown in figure 1. An original movie file is pass into the model as an input, where the full movie is broken into several movie frames. The frames are then forwarded to the first layer of the model which is made up of Long Short-Term Memory(LSTM), which handles the exploitation of intra-subshot temporal dependency within the movie frames. The second layer is abidirectional LSTM with a forward and backward inter-subshottemporal dependency, and the output obtained from the second layer is used for the prediction of certain subshot being valuable to be selected into the summary.

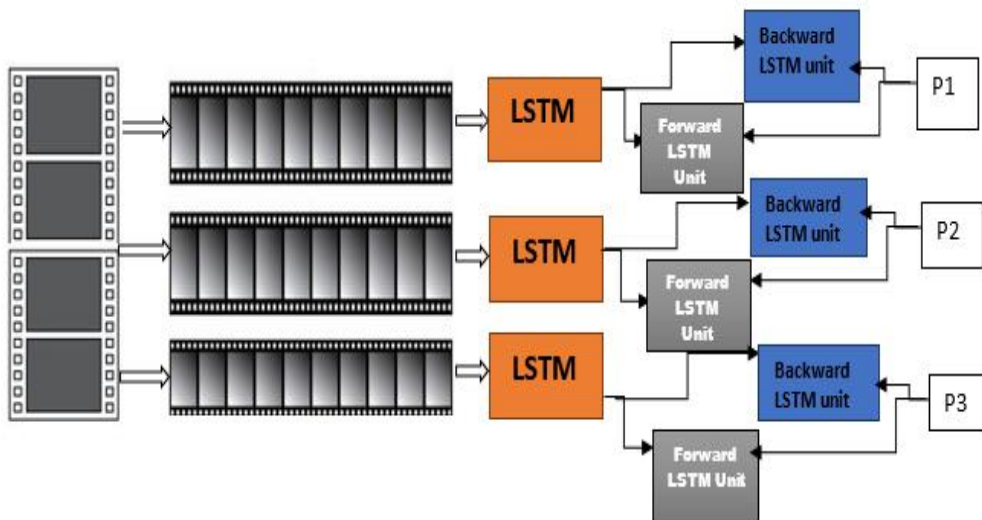


Fig 1: Architecture of the Proposed framework

4. FRAMEWORK DESIGN

The purpose of this framework design is to create a technical solution that satisfies the functional requirements for automatic movie sequences summarization. This system is design using Python 3.9. The implementation was done using Keras and TensorFlow frameworks. The audio sound

processing and concatenating selected keyframe is handled by Fast Forward Moving Picture Expert Group (FFMPEG). The system is designed to accept input movie file and split them into video (sequence) and sound (sound data), keyframes extracted, and a comprehensive and precise movie summary obtain.

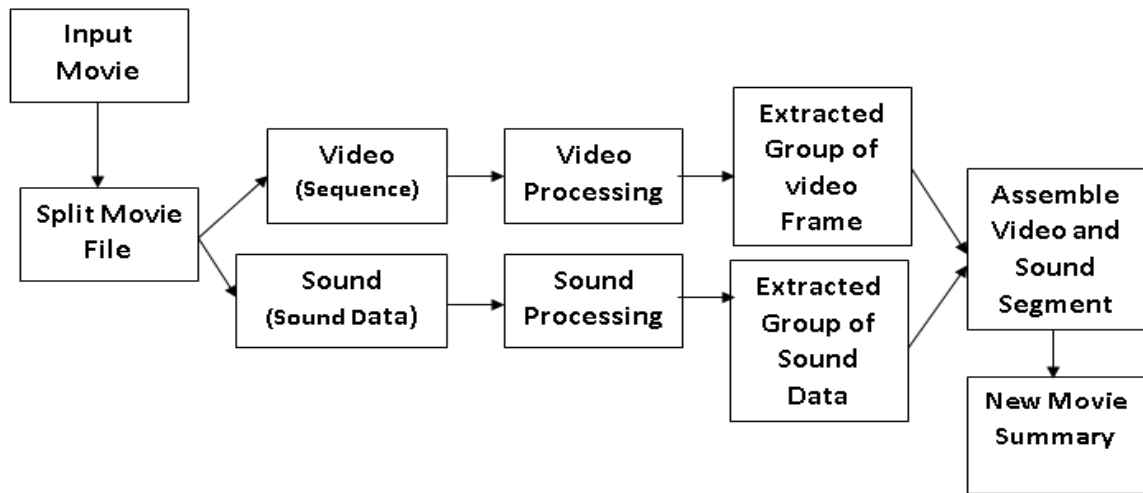


Fig 2: Framework for Automatic audio sound and video sequence Summarization

4.1 Summarization

The proposed framework in figure 2 can be demonstrated in four steps: the first step is to split the input movie file into sound and video data. The second step is to process both the video data through frame extraction and selecting the desired group of frames based on threshold value. Third step is the sound processing that is obtainable by analyzing audio data

and getting the sound data segment selected relative to the group of frames selected by the RNN model. The final step is the assemble process to construct a new video summary.

4.2 RNN Model framework

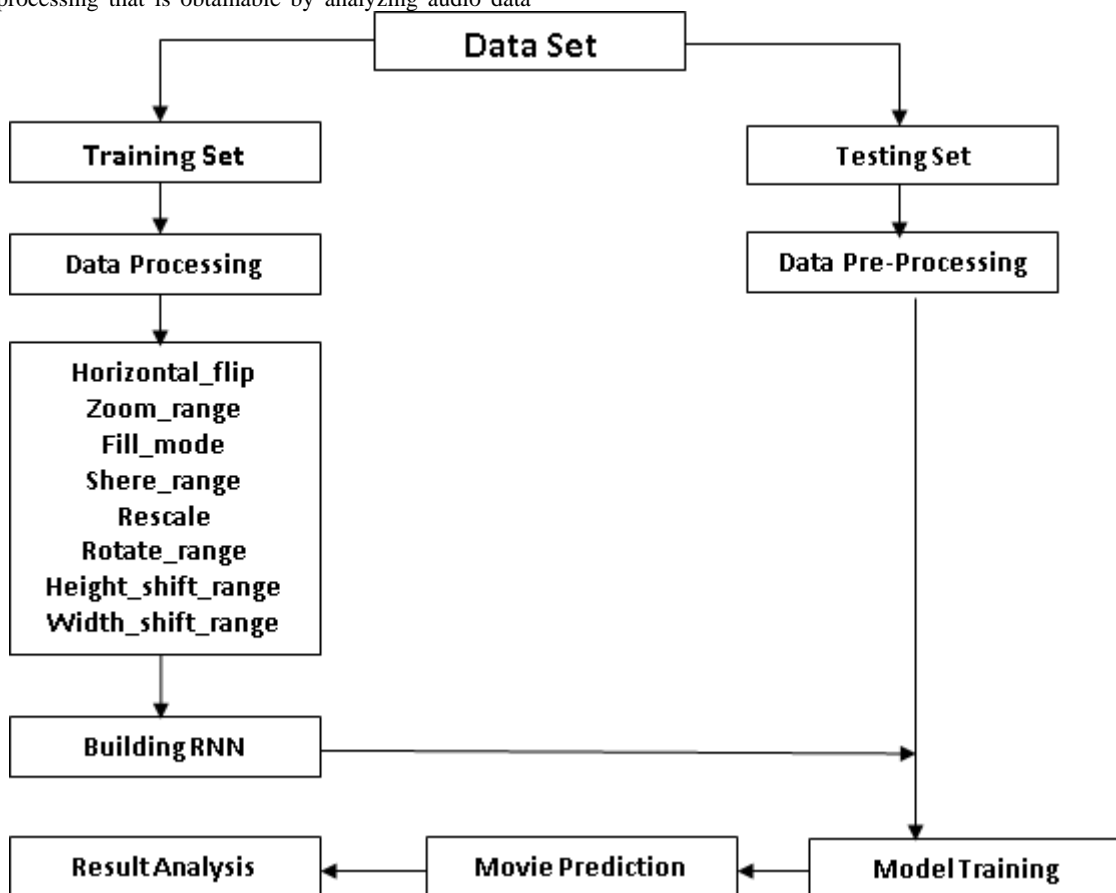


Fig 3: Framework of Building the RNN Model

Figure 3 is the construction of the model and the framework of the model showing two sets of data from the dataset, the train set and test set. This model shall use two benchmark dataset TvSum [3] and SumMe [4] for training the model for effective prediction. The train set goes through augmentation process to

build the model, while the testing dataset is processed directly to the trained model for prediction.

To address the peculiar problem of Nollywood movie sequences and audio sound data summarization, an embedded custom dataset of Nollywood movie was created

and added to TvSum. The essence of creating a custom dataset is to enable the model gain more weight patterning to Nollywood movie scenes during training to foster accurate predictions of the various movie genre produced by the movie industry.

5. NOLLYWOOD MOVIE SUMMARIZATION USER INTERFACE

Figure 4.1 depicts the user interface of the proposed movie sequence summarization system. The user interface assists the users to obtain a comprehensive but precise movie summary

and to give a preferred summary length for a movie. The user selects an original movie file into the model by selecting it from a movie folder. The new movie summarization system input is a movie file with full content as the original. The target output is a small content of important keyframe from the original input movie to produce a summary that expresses without having to watch the full content and without losing interesting information.

The interface displays the summarization progress as well as the total time taken for the entire process to be completed. The movie is summarized and kept in the originated movie folder.

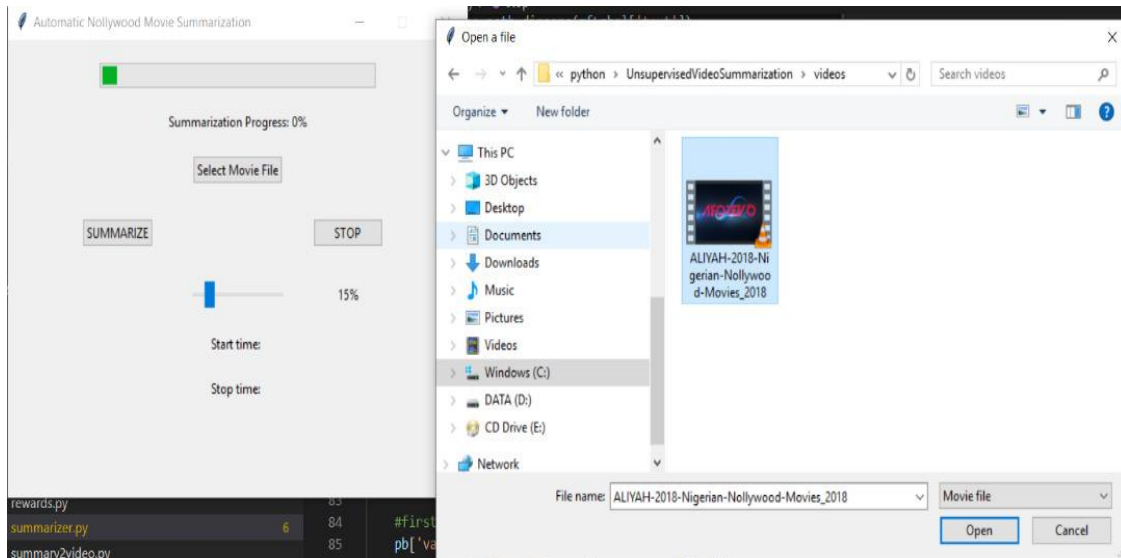


Fig 4.1: Input interface

Figure 4.1 is the main menu of the automatic Nollywood movie summarization user interface for the implementation of the proposed system showing how a movie file can be selected for

summarization and how the expected summarized percentage can be specified

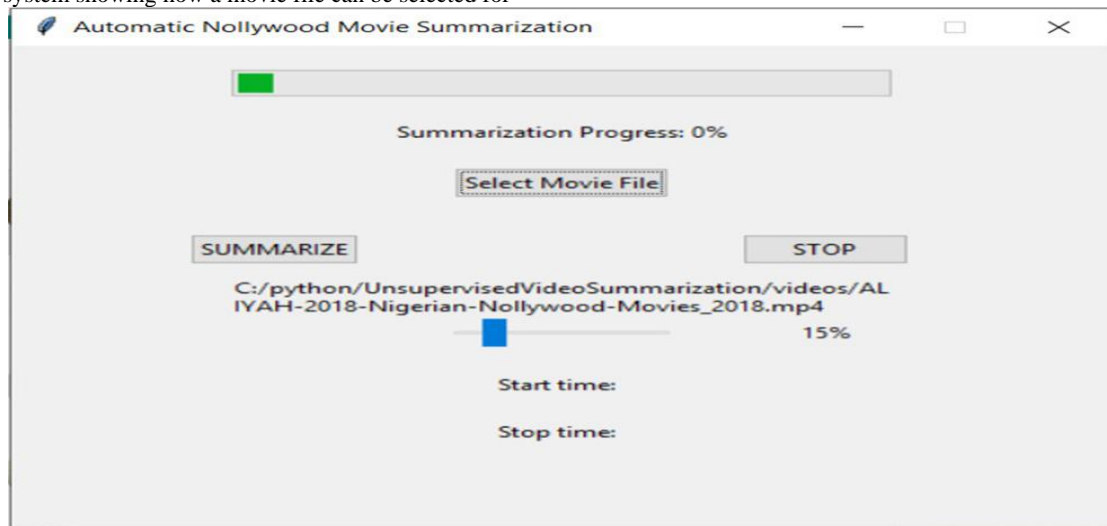


Fig4.2: User interface showing selected movie file.

Figure 4.2 is the automatic Nollywood movie summarization user interface for the implementation of the proposed system

showing a selected movie file selected for summarization with summarization progress detail.

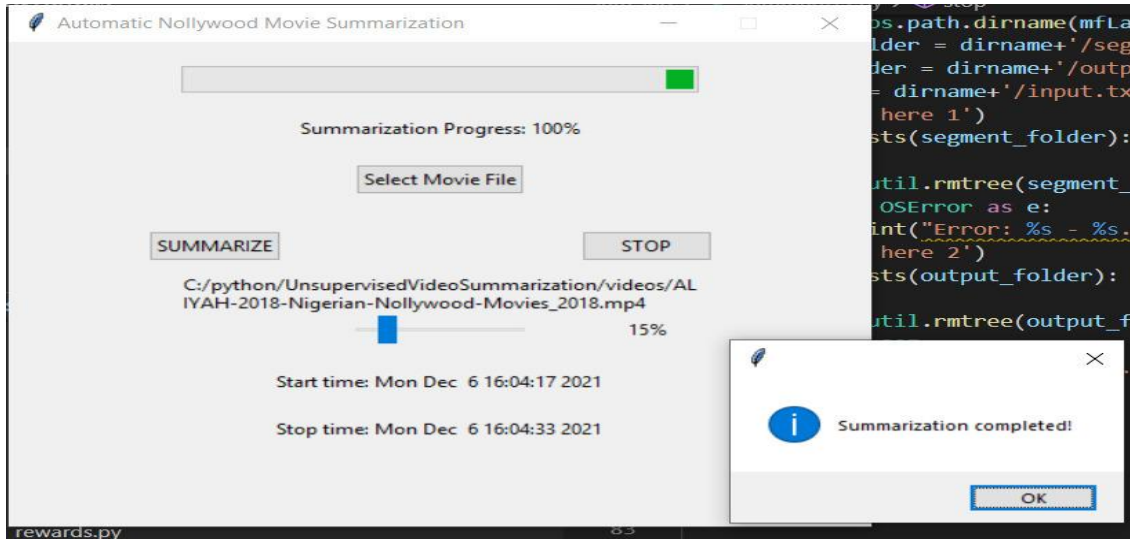


Fig 4.3: User interface showing completed summarization process.

Figure 4.3 is the automatic Nollywood movie summarization user interface for the implementation of the proposed system

showing a successfully completed summarization process, displaying the start time and date for the summarization.

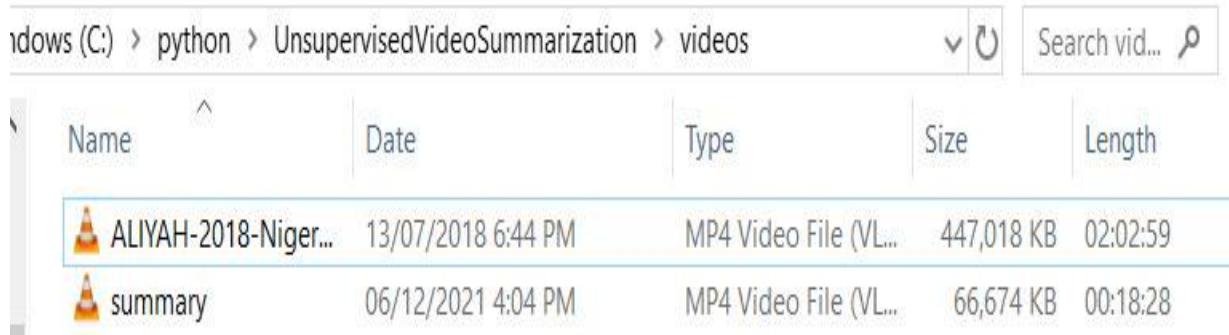


Fig4.4: folder containing the original movie and summary

Figure 4.4 is the automatic Nollywood movie summarization user folder showing the output of a completed summarization

process with the file size and movie length of summary smaller than the original movie.

Table 5.4 Results analysis obtained from summarization of different movies

Movie Name	Original movie size (kb)	Summary size (kb)	Original movie length	Duration (seconds)	Percentage of summary (%)	Frames	Extracted frames
Anchor Baby	324,810	46,433	01:35:01	13	15%	6958	139
Baby Dady	195,538	26,756	1:02:33	9	15%	5445	92
Daughter in-law	246,661	33,403	02:53:48	11	15%	1490	97
Devil in the Church	285,755	39,724	02:38:20	10	15%	11,658	94
Aliyah	447,018	22,069	02:02:59	14	5%	9,247	180

In order to analyze if the proposed system works properly, a sample of 5 Nollywood movies was chosen, all in MP4 format (30 fps, 320 x 240 pixels), pertaining to romance and drama as contained in table 5.4. The experiment displays the percentage of the summaries according to the user's preference. The total number frames contained in the movies as well as the extracted frames that make up the summary.

The duration of individual movies varies from 9 to 14 seconds. This experiment was carried out on an HP OMEN Intel core i7 2.20 GHz 2.21 GHz with 16GB RAM.

6. SUMMARY AND RECOMMENDATIONS

The goal of this summarizer is to reduce viewing time, make the process of choice movie to watch easier, less biased than humans and present information in a way that is attractive to the viewers and conserve the main content of the original movie. This novel representation will not only save processing time of Nollywood movie trailers production but will also save storage space as well as produce a high-quality multimedia advertisement medium for the home movie industry.

The researcher wishes to make some recommendations, which, if taken into consideration, might bring some positive changes to the current approach.

Nollywood industry and home movie lovers should:

- Abandon the manual approach of movie trailer production and adopt the new state of the art multimedia approach as an alternative.
- The National Film and Video Censors Board (NFVCB) should immediately adopt the new system for classifying all Nollywood movies produced.
- Local TV stations in the country should embrace the new system for producing summaries of its daily broadcast.
- Stakeholders in the movie industry like producers and directors of movie should work closely with researchers in the area of computer vision to provide accurate information on specific contents on movie scenes making. This would ensure that the researcher does not have to fabricate information from what is implied.
- Break new grounds by stepping up to use of modern sophisticated software for digital advertisement.
- Researchers in this area should also endeavor to improve on the custom dataset for Nollywood movie industry that has already created.

7. REFERENCES

- [1] Chowdhury Mridul, Landesz Tama, Santini Massimiliano, Tejada Luis, & Visconti Gloria. 2008. Nollywood: The Nigerian Film Industry. Harvard Business School Publishers.
- [2] Stephen, G. 2012. The Art of First Impressions: How to Cut a Movie Trailer. film maker magazine. <http://filmmakermagazine.com>.
- [3] Yale Song, Jordi Vallmitjana, Amanda Stent, and Alejandro Jaimes. 2015. TVSum: summarizing web videos using titles, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 5179–5187, 2015.
- [4] Gygli, M.; Grabner, H.; Riemenschneider, H.; and Van Gool, L. 2014. Creating summaries from user videos. In ECCV, 505–520. Springer.
- [5] Sandra Eliza Fontes de Avila; Ana Paula Brandão Lopes; Antonio da Luz Jr.; Arnaldo de Albuquerque Araújo. 2011. VSUMM: A mechanism designed to produce static video summaries and a novel evaluation method. , 32(1), 56–68. doi:10.1016/j.patrec.2010.08.004.
- [6] Sanjay K. Kuanar, Rameswar Panda, and Ananda. S. Chowdhury. 2013. Video key frame extraction through dynamic delaunay clustering with a structural constraint, J. Vis. Commun. Image Represent., vol. 24, no. 7, pp. 1212–1227.
- [7] Shaohui Mei, Genliang Guan, Zhiyong Wang, Mingyi He, and David Dagan Feng. 2015. Video summarization via minimum sparse reconstruction, Pattern Recognition, vol. 48, no. 2, pp. 522–533.
- [8] Yang Cong, Junsong Yuan, and Jiebo Luo. 2012. Towards scalable summarization of consumer videos via sparse dictionary selection, IEEE Trans. Multimedia, vol. 14, no. 1, pp. 66–75,.
- [9] Yanwei Fu, Yanwen Guo, Yanshu Zhu, Feng Liu, Chuanming Song, and Zhihua Zhou. 2010. Multi-View Video Summarization, IEEE Trans. Multimedia, vol. 12, no. 7, pp. 717–729,.
- [10] Meng Wang, Richang Hong, Guangda Li, Zhengjun Zha, and Shuicheng Yan. 2012. Event driven web video summarization by tag localization and keyshot identification, IEEE Trans. Multimedia, vol. 14, no. 4, pp. 975–985.
- [11] Potapov, D.; Douze, M.; Harchaoui, Z.; and Schmid, C. 2014. Category-specific video summarization. In ECCV, 540–555. Springer.
- [12] Michael Gygli, Helmut Grabner, Hayko Riemenschneider, and Luc Van Gool. 2014. Creating summaries from user videos. In ECCV - 13th European Conference on Computer Vision. 2014, vol. 8695, pp. 505–520, Springer.
- [13] Hussain, T; Muhammad, K; Ullah, A; Cao, Z; Baik, Sung W; de Albuquerque, Victor C. 2019. Cloud-Assisted Multi-View Video Summarization using CNN and Bi-Directional LSTM. IEEE Transactions on Industrial Informatics. 1 1. doi:10.1109/TII.2019.2929228.
- [14] S. Khan, T. Yairi. 2018. A review on the application of deep learning in system health management. Mechanical Systems and Signal Processing 107 241–265.
- [15] O.O Tumenay, F. S. Bakpo, D. U. Ebem 2022. Nollywood Movie Sequences Summarization Using a Recurrent Neural Model. Quest Journal of Software Engineering and Simulation Volume 8 ~ Issue 1 (2022) pp: 06-11 ISSN(Online) :2321-3795 ISSN (Print):2321-3809.