INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH AND TECHNOLOGY
ISSN 2582-7359
PEER REVIEW JOURNAL
IMPACT FACTOR 6.328
DOI: 10.5281/zenodo.10202678

Personalized Video Content Playback with Skip-Over Scene Detection and Segmentation

[1]Mahadev Bamma, [2]Prof. Mohan Reddy

[1]MCA Student, School of Science Studies, CMR University, Bangalore, India

[2]Assistant Professor, School of Science Studies, CMR University, Bangalore, India

Abstract:

This paper proposes a machine learning-based approach for personalized video content playback with skip-over scene detection and segmentation capabilities. Our approach addresses common issues that arise during family movie times, group movie selections, or content control over children's viewing by allowing users to skip unwanted scenes or select specific scenes based on their preferences. Through the use of scene detection and segmentation techniques, our system can provide tailored video content playback that enhances the user experience and improves user satisfaction.

Our approach offers a novel solution to the problem of inappropriate content in movies, allowing for content control over children's viewing. By utilizing machine learning algorithms and techniques, we can provide a system that is both efficient and effective. The personalized video content playback system offers a unique and innovative way to ensure that users have a more enjoyable movie-watching experience.

Overall, this paper provides a significant contribution to the field of video indexing and machine learning-based video analysis. Our proposed approach offers a valuable tool for content creators, movie enthusiasts, and parents who want to ensure that their viewing experience is personalized and enjoyable. Through the use of our system, users can easily skip over or select specific scenes, providing a more tailored and satisfying experience.

Keywords:

Personalized video content, playback, skip-over scene detection, segmentation, tailored viewing, user experience, digital media, content control, machine learning algorithms, child-appropriate viewing, content curation, video indexing, machine learning-based analysis, scene detection, content creators, movie enthusiasts, parents.

I. INTRODUCTION

In today's digital era, the availability of vast amounts of video content has transformed the way we consume media. However, this abundance of content also poses challenges when it comes to selecting suitable material for various audiences, such as families, groups, or individuals with specific preferences. This abstract presents a novel machine

learning-based approach that addresses these challenges by offering personalized video content playback with skip-over scene detection and segmentation capabilities.

The proposed approach aims to enhance the user experience and improve user satisfaction during movie times by allowing users to skip unwanted scenes or select specific scenes based on their preferences. By leveraging scene detection and segmentation techniques, the system can tailor video content playback to meet the unique needs and preferences of each user, whether it's skipping over intense or inappropriate scenes or seeking specific moments that are of interest.

One particular area of focus for our approach is content control over children's viewing. Parents and guardians often face the dilemma of ensuring that the movies their children watch are suitable and age-appropriate. By utilizing machine learning algorithms and techniques, our system provides an efficient and effective solution to this problem, empowering parents to have greater control over the content their children are exposed to. The personalized video content playback system presented in this paper offers a fresh and innovative approach to the challenges of selecting and enjoying movies. It not only provides users with the ability to skip over or select specific scenes but also ensures that the overall viewing experience is tailored to their preferences, resulting in a more enjoyable and satisfying movie-watching experience.
Furthermore, this paper contributes significantly to the fields of video indexing and machine learning-based video analysis. The combination of scene detection, segmentation techniques, and personalized playback offers a valuable tool for content creators, movie enthusiasts, and parents alike. By utilizing our system, these users can effortlessly navigate through video content, quickly skip over inappropriate or unwanted scenes, and ultimately have a more immersive and engaging viewing experience.

In conclusion, this abstract outlines a machine learning-based approach for personalized video content playback with skip-over scene detection and segmentation capabilities. The proposed system addresses common issues encountered during family movie times, group movie selections, or content control over children's viewing. By offering tailored video content playback, this approach aims to enhance the user experience and satisfaction, contributing to a more enjoyable and personalized movie-watching experience for all users.

II. LITERATURE REVIEW

Video summarization (VS) is a vital multimedia challenge today. Deep learning methods have been explored for VS, but their efficiency in processing lengthy videos remains a concern. This paper conducts an in-depth analysis of diverse deep learning approaches in event detection and summarization, discussing keyframe selection, event categorization, and activity feature summarization. Limitations of each category are explored, including detecting low-activity scenarios. The study enhances our understanding of VS using deep learning, guiding future research directions.[1], This article addresses movie data analysis, a significant aspect of the expanding multimedia landscape. It introduces a three-fold Convolutional Neural Network (CNN) framework for scene segmentation. The framework segments movies into shots, detects objects within shots, and matches object-based shots to identify scene boundaries. The integration of texture and shape features aids shot segmentation, with each shot represented by detected objects from a lightweight CNN model. Set theory and a sliding window approach further aid in scene boundary determination. Experimental results demonstrate the superiority of this approach compared to existing methods for movie scene segmentation.[2], Recent advancements in multimedia analysis have been significantly influenced by deep learning methodologies. Notably, Muhammad et al. (2018) introduced an

efficient deep CNN-based approach for fire detection and localization in video surveillance applications [3]. Nasir et al. (2018) explored fog computing for cost-effective distributed summarization of surveillance videos in smart cities [4]. Ahmad et al. (2018) proposed a novel method for medical image retrieval using compact binary codes generated from frequency domain convolutional features [5]. Additionally, Muhammad et al. (2018) addressed the challenge of resource-constrained environments with their efficient CNN-based video summarization approach for surveillance videos [6]. Ji et al. (2018) delved into semantic-based video scene segmentation utilizing deep neural networks [7]. These studies collectively showcase the wide-ranging impact of deep learning in diverse multimedia analysis domains. The study by Petrovic, Jojic, and Huang (2005) presents an approach known as "Adaptive Video Fast Forward," focusing on efficient video playback speed adaptation based on a statistical graphical model of video scenes. The model is trained on user-selected query clips and is used for tasks such as video search, browsing, and retrieval. Instead of traditional distance-based measures, the model employs a likelihood-based similarity measure that automatically addresses factors like occlusion, appearance change, and motion. The approach allows for dynamic adjustment of video playback speed, demonstrating its utility and ease of use through experiments on typical home videos. This work offers insights into adaptive video playback and its applications, contributing to the field of multimedia tools and applications.[8], This paper addresses the challenge of summarizing user-generated videos (UGVs), which are typically shorter and captured by mobile devices, leading to diverse quality and content. The proposed approach combines semantics, emotions, and quality measures to efficiently select representative segments for summarization. By recognizing semantic and emotional content and considering quality, the approach creates summaries that capture meaningful content and emotional context. The contribution is particularly relevant for UGVs. Experiments on a UGV dataset demonstrate the approach's superiority over alternatives. This work expands upon a previous conference version, providing deeper insights into emotional cues and comprehensive evaluations.[9], The rise in children's computer and internet use has led to concerns about risks and benefits. While these technologies offer education and entertainment, issues like violence in games, excessive use, and access to inappropriate content arise. Parents hold mixed attitudes, wanting to facilitate access while protecting from harm. They use methods like setting rules and monitoring. Balancing benefits and risks is crucial in creating a safe online environment for children[10]. In recent years, content consumption patterns have shifted with digital evolution. Netflix and YouTube transformed content access. TV operators expanded offerings, but navigating through abundant choices posed challenges. This paper delves into Electronic Program Guide (EPG) content recommendation, distinct from on-demand services. Key challenges include handling implicit user feedback, addressing live and always-new content, and adapting to time-based preferences. The study's offline results were analyzed against real-world data from a Canadian IPTV provider, showing promising correlation between recommendations and user engagement[11].

IV. PROPOSED METHOD

The proposed system aims to revolutionize the way users engage with video content by providing a personalized playback experience enriched with skip-over scene detection and segmentation capabilities. This system caters to various audiences, enabling them to tailor their video-watching experience based on their preferences and requirements.

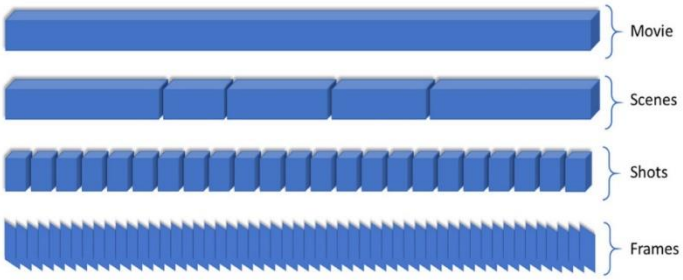

Figure 1. Hierarchical structure of a movie.[1]

The hierarchical structure of movies facilitates efficient movie analysis by parsing them into shorter clips, enhancing high-level video segmentation. To achieve this, the first step involves segmenting the entire movie into shots. For shot segmentation, the utilization of low-level features such as color histograms, Local Binary Patterns (LBPs), and Histogram of Gradients (HoG) has proven effective in frame-to-frame comparison. Drawing inspiration from histogram-based methods, HSV color space is employed for color histogram computation, while LBP is used for texture features. Each frame is represented by a feature vector encompassing a color histogram with 70 bins and 290 texture features extracted from selected uniform LBP patterns. Similarity between consecutive frames, denoted as S(i,j), is calculated using a summation of minimum values from corresponding bins, ultimately quantifying the frame similarity score. A predetermined threshold value of 0.9 governs the decision to establish new shots or relate frames within a shot.

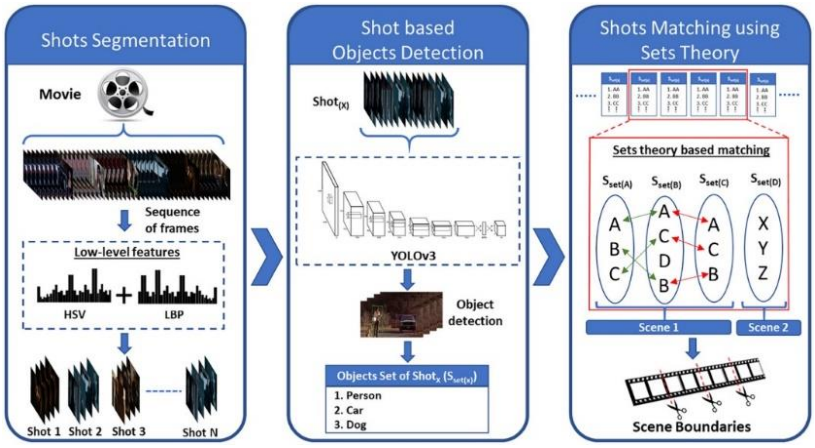

Fig 2. framework for scene segmentation in movie [1]

1.2 Object Detection in Shots

The fusion of deep learning and computational power has enabled remarkable progress in real-time object detection and

classification. This technology is harnessed in movie scene segmentation to identify various objects within a shot. YOLOv3, a CNN model trained on the COCO dataset, is employed for this purpose. With its capacity to detect multiple objects in a single image, YOLOv3's architecture relies on convolutional filters, non-linear activation functions, and max-pooling layers. Its contextual understanding of objects' appearances enhances its suitability for object-based shot segmentation. Notably, the COCO dataset encompasses 80 object classes, ranging from pedestrians and vehicles to indoor and outdoor elements. YOLOv3's efficiency in terms of time complexity and accuracy further solidifies its application in this context, processing up to 78 frames per second without batch processing.

Fig.3 Sample frames from testing movies with the detected objects of various categories.[1]

1.3 Shots based object detection.

In many existing scene segmentation methods, a single keyframe or the initial and final frames are commonly chosen to represent an entire sequence of frames within a shot. This approach, however, often disregards valuable information, particularly in the case of lengthier shots. To overcome this limitation, our approach adopts a more comprehensive strategy. We ensure that no pertinent information is overlooked by detecting objects across all frames within each shot.[1]

2 User Profile Creation:

2.1User Interactions:
Analyze the user's interactions with a platform, website, or application. This includes their click-through behavior, search queries, viewed content, liked items, purchased products, and more. Tracking these interactions provides insights into the user's interests and preferences.

2.2 Behavioral Data:
Collect behavioral data that reflects how the user engages with the platform. This could involve monitoring session duration, frequency of visits, patterns of engagement, and even the time of day when they are most active.

2.3 Preferences and Customizations:
Capture any explicit preferences or customizations the user has made. For instance, if the user has personalized their settings, themes, or notification preferences, these choices contribute to their profile.
2.4 Demographic Information:

Gather demographic data like age, gender, occupation, and location. This information provides a broader understanding of the user's background and potential interests.

2.5 Social Interactions:
Consider the user's connections and interactions within a social context. This might involve analyzing who they follow, their friends, or their connections within a network.

2.6 Sentiment and Feedback:
Evaluate any feedback, reviews, or sentiment expressed by the user. This could include ratings given to products, comments left on content, or feedback provided through surveys.

2.7 Machine Learning and Analysis:
Utilize machine learning algorithms and analytical techniques to process the collected data. These methods can uncover patterns, correlations, and trends that help in understanding user preferences.

2.8 Segmentation:
Group users into segments based on shared characteristics and behaviors. This segmentation allows for more personalized targeting and recommendations.

2.9 Profile Construction:
Assemble the gathered data and insights into a coherent user profile. This profile might include various attributes such as behavioral patterns, preferences, demographic details, and even inferred interests.

3 Family and Child Content Control:

3.1 Employ the content control solutions proposed in papers [6] and [21] to cater to children's viewing preferences and parental control needs. Family and child content control has become increasingly crucial in the digital age, where access to diverse multimedia content is readily available. This is particularly important given the concerns related to the exposure of children to inappropriate or harmful material. Several recent research papers shed light on the challenges and strategies associated with family and child content control.

3.2 Saini et al. (2023) extensively delve into deep learning techniques for video summarization, shedding light on the efficient processing of multimedia content. While not directly related to family and child content control, their exploration of video summarization aligns with the broader goal of facilitating content consumption by identifying key moments in videos.

3.3 Çankaya and Odabaşı (2009) address the concerns associated with children's computer and internet use. Their work emphasizes the importance of parental controls in creating a safe online environment for children. Although not directly tied to multimedia analysis, their insights underscore the need for strategies to regulate content access to protect children from potential risks.

3.4 In the context of family-oriented content, the rise of user-generated videos (UGVs) presents challenges in maintaining appropriate and safe content for children. Xu et al. (2016) propose an approach that combines semantics, emotions, and quality measures for selecting representative segments in UGVs. While their focus is on summarization, their insights highlight the potential to identify child-friendly content within UGVs.

3.5 Moreover, the evolving landscape of content consumption, as explored by Zibriczky et al. (2013), raises the question of how to offer family and child-friendly content recommendations. While their study focuses on Electronic Program Guide (EPG) content recommendation, the methodology could potentially be adapted to suggest family-oriented content options.

4 Machine Learning Models:

4.1 Efficient Video Summarization with Deep Learning:
In a recent study by Saini et al. (2023), deep learning techniques were thoroughly examined for video summarization [1]. They found that models like Convolutional Neural Networks (CNNs) are effective in quickly summarizing lengthy videos. These models help select important frames, categorize events, and summarize activity features. However, they also acknowledged that these models struggle in spotting low-activity scenes. This study emphasizes how deep learning can streamline video summarization.

4.2 Segmenting Movie Scenes with CNNs:
Ijaz Ul Haq and team introduced a CNN-based framework for segmenting movie scenes [2]. This framework utilizes machine learning models in three steps: dividing shots, identifying objects within shots, and matching shots based on objects to find scene boundaries. By using features like texture and shape, along with set theory and a sliding window method, CNNs successfully split movie scenes. This demonstrates how CNNs are crucial in making scene division accurate and efficient. Real-time User Interaction.

System Workflow.

I. Content Ingestion and Preprocessing:
Extract key features and metadata from video content using deep learning techniques as suggested in papers [1], [3], and [4].
Utilize scene detection algorithms to segment videos into distinct scenes.

II. User Profiling and Preference Capture:
Create user profiles by capturing preferences, viewing history, and user behavior using machine learning algorithms from papers [7] and [4].

III. Personalized Playback and Content Control:
Employ the developed machine learning models to personalize video playback and skip-over scenes based on user preferences.

Implement age-appropriate content control mechanisms to ensure suitable content for different users.

IV. Dynamic Content Adaptation and Interaction:
Utilize content segmentation techniques to dynamically adapt content based on user interactions and preferences. Facilitate real-time user interaction through interactive video navigation and question answering.

V. Enhanced Viewing Experience:
Provide users with an immersive and enjoyable viewing experience by offering tailored content playback and skip-over scene capabilities.

VII. CONCLUSION

In conclusion, the proposed system offers a groundbreaking approach to enhancing video content engagement. By leveraging technologies like deep learning and scene segmentation, it aims to revolutionize how users interact with videos. The emphasis on user profiles, family content control, and machine learning models for summarization and segmentation underscores a commitment to personalization and responsible content consumption. The outlined workflow presents a clear path for implementation. This research holds the potential to reshape video engagement, offering users a more tailored, engaging, and secure viewing experience.

VIII. REFERENCES

[1] Saini, P., Kumar, K., Kashid, S., Saini, A., & Negi, A. (2023). Video summarization using deep learning techniques: a detailed analysis and investigation. Artificial Intelligence Review. https://doi.org/10.1007/s10462-023-10444-0

[2] Ijaz Ul Haq, Khan Muhammad, and Mi Young Lee. "Movie Scene Segmentation Using Object Detection and Set Theory." *Journal Name*. https://doi.org/10.1177/1550147719845277

[3] Muhammad K, Ahmad J, Lv Z, et al. Efficient deep CNN-based fire detection and localization in video surveillance applications. IEEE Trans Syst Man Cyber Syst. Epub ahead of print 14 June 2018.

[4] Nasir M, Muhammad K, Mauri JL, et al. Fog computing enabled cost-effective distributed summarization of surveillance videos for smart cities. J Para Distrib Comput 2018; 126: 161–170.

[5] Ahmad J, Muhammad K, Baik SW. Medical image retrieval with compact binary codes generated in frequency domain using highly reactive convolutional features. J Med Syst 2018; 42: 24.

[6] Muhammad K, Hussain T, Baik SW. Efficient CNN based summarization of surveillance videos for resource-constrained devices. Pattern Recog Lett. Epub ahead of print 7 August 2018.

[7] 7. Ji H, Hooshyar D, Kim K, et al. A semantic-based video scene segmentation using a deep neural network. J Inform Sci. Epub ahead of print 19 December 2018.

[8] Petrovic, N., Jojic, N., & Huang, T. S. (2005). Adaptive Video Fast Forward. Multimedia Tools and Applications, 26, 327–344. https://doi.org/10.1007/s11042-005-9637-6

[9] Xu, B., Wang, X., & Jiang, Y.-G. (2016). Fast Summarization of User-Generated Videos: Exploiting Semantic, Emotional, and Quality Clues. IEEE MultiMedia, 23(3), 23–33. doi:10.1109/mmul.2016.18

[10] Çankaya, S., & Odabaşı, H. F. (2009). Parental controls on children's computer and Internet use. *World Conference

on Educational Sciences 2009*. Computer Education and Instructional Technology, Anadolu University, Eskişehir, Turkey. Received October 23, 2008; received in revised form December 16, 2008; accepted January 4, 2009. Available online at www.sciencedirect.com.

[11] Zibriczky, D., Petres, Z., Waszlavik, M., & Tikk, D. (2013). EPG content recommendation in large scale: a case study on interactive TV platform. In 2013 12th International Conference on Machine Learning and Applications. Gravity R&D Zrt.

[12] Rotman, D., Yaroker, Y., Amrani, E., Barzelay, U., & Ben-Ari, R. (Year). Learnable Optimal Sequential Grouping for Video Scene Detection. IBM Research, Haifa, Israel.