information, offering a multi-modal perspective for content analysis.

We employ the YoloV8m deep learning model to perform object detection and extraction frame by frame, allowing us to select crucial objects based on their frequency within the video content. Additionally, we utilize STT to transcribe audio content into textual scripts, providing a text-based representation of the video's audio component. These two elements, object frequency analysis and textual scripts, are seamlessly integrated on a frame-by-frame basis, creating rich contextual data that captures both visual and auditory elements. This contextual data forms the foundation for our innovative video clipping algorithm, optimizing the categorization and segmentation of video content. Furthermore, our model supports fine-tuning through cross-validation, ensuring script accuracy and optimizing model hyperparameters. By expanding the metadata associated with video content, our approach significantly enhances video search and recommendation systems. The enriched metadata provides a comprehensive dataset for analysis, leading to improved content recommendations.

Moreover, our method streamlines video content management, particularly for lengthy videos, live streams, and edited content, offering a more efficient solution for content creators and viewers alike. Overall, our proposed approach has the potential to revolutionize the way video content is managed, curated, and presented across various platforms. It empowers content creators and viewers with a more efficient, nuanced, and personalized video experience, ultimately advancing the capabilities of video service platforms in the rapidly evolving multimedia landscape.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2020R1A6A1A03040583). Additionally, this work was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(No: RS-2023-00248899).

REFERENCES

[1] Kim J. C. and Chung K. Y., "Knowledge expansion of metadata using script mining analysis in multimedia recommendation", *Multimedia Tools and Applications*, vol. 80, pp. 34679-34695, 2020.

[2] Lee J. H., "The Growth and Impact of OTT on Video Viewing Behavior", *Asian-pacific Journal of Convergent Research Interchange*, vol. 6, no. 1, pp. 41-50, 2020.

[3] Shah S. and Mehta N., "Over-the-top (OTT) streaming services: studying users' behaviour through the UTAUT model", *Management and Labour Studies*, vol. 48, no. 4, pp. 531-547, 2023.

[4] Luo, M., Chen, F., Cheng, P., Dong, Z., He, X., Feng, J. and Li, Z., "Metaselector: Meta-learning for recommendation with user-level adaptive model selection", *In Proceedings of The Web Conference 2020*, pp. 2507-2513, 2020.

[5] N. Silva, T. Silva, H. Werneck, L. Rocha and A. Pereira, "User cold-start problem in multi-armed bandits: When the first recommendations guide the user's experience", *ACM Transactions on Recommender Systems*, vol. 1, no. 1, pp. 1-24, 2023.

[6] Hao, B., Yin, H., Zhang, J., Li, C. and Chen, H., "A Multi-strategy-based Pre-training Method for Cold-start Recommendation", *ACM Transactions on Information Systems*, vol. 41, no. 2, pp. 1-24, 2023.

[7] A. Ishikawa, E. Bollis and S. Avila, "Combating the elsagate phenomenon: Deep learning architectures for disturbing cartoons," *in Proc.IWBF'19*, pp. 1-6. 2019.

[8] Matakupan, N. I., "The Study of'Don't Hug Me I'm Scared'Web Series Storytelling For IP Design Regarding Safe Viewing Content For Children", *IJVCDC (Indonesian Journal of Visual Culture, Design, and Cinema)*, vol. 2, no. 2, pp. 172-177, 2023.

[9] Yousaf, K. and Nawaz, T., "An attention mechanism-based CNN-BiLSTM classification model for detection of inappropriate content in cartoon videos", *Multimedia Tools and Applications*, pp. 1-24, 2023.

[10] Huang, S., Liu, G., Chen, Y., Zhou, H. and Wang, Y., "Video Recommendation Method Based on Deep Learning of Group Evaluation Behavior Sequences", *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 37, no. 02, pp. 2352002, 2023.

[11] A. Yousaf., A. Mishra., B. Taheri and M. Kesgin, "A cross-country analysis of the determinants of customer recommendation intentions for over-the-top (OTT) platforms," *Information & Management*, vol. 58, no. 8, pp. 103543, 2021.

[12] Hashemi, M., "Web page classification: a survey of perspectives, gaps, and future directions", *Multimedia Tools and Applications*, vol. 79, no. 17-18, pp. 11921-11945, 2020.

[13] Hesmondhalgh, D. and Lotz, A., "Video screen interfaces as new sites of media circulation power", *International Journal of Communication,* vol. 14, pp. 386-409, 2020.

[14] Patnaik, R., Shah, R. and More, U., "Rise of OTT platforms: effect of the C-19 pandemic", *PalArch's Journal of Archaeology of Egypt/Egyptology*, vol. 18, no. 7, pp. 2277-2287, 2021.

[15] Singh, N., Arora, S. and Kapur, B., "Trends in over the top (OTT) research: a bibliometric analysis", *VINE Journal of Information and Knowledge Management Systems*, vol. 52, no. 3, 411-425, 2022.

[16] Sontakke, K. S., "Trends in OTT Platforms Usage During COVID-19 Lockdown in India", *Journal of Scientific Research*, vol. 65. no. 8, pp. 23, 2021.

[17] Sun, C., Jia, Y., Hu, Y. and Wu, Y., "Scene-aware context reasoning for unsupervised abnormal event detection in videos", *In Proceedings of the 28th ACM International Conference on Multimedia*, pp. 184-192, 2020.

[18] Ramachandra, B., Jones, M. J. and Vatsavai, R. R., "A survey of single-scene video anomaly detection", *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no.5, pp. 2293-2312, 2020.

[19] Raja, R., Sharma, P. C., Mahmood, M. R. and Saini, D. K., "Analysis of anomaly detection in surveillance video: recent trends and future vision", *Multimedia Tools and Applications*, vol. 82, no. 8, pp. 12635-12651, 2023.

[20] Luo, H., Ji, L., Zhong, M., Chen, Y., Lei, W., Duan, N. And Li, T., "Clip4clip: An empirical study of clip for end to end video clip retrieval and captioning", *Neurocomputing*, vol. 508, pp. 293-304, 2022.

[21] Haq, H. B. U., Asif, M. and Ahmad, M. B., "Video summarization techniques: a review", *Int. J. Sci. Technol. Res*, vol. 9, no. 11, pp. 146-153, 2020.

[22] Workie, A., Sharma, R. and Chung, Y. K., "Digital video summarization techniques: A survey", *Int. J. Eng. Technol*, vol. 9. no. 1, pp. 81-85, 2020.

[23] Tiwari, V. and Bhatnagar, C., "A survey of recent work on video summarization: approaches and techniques", *Multimedia Tools and Applications*, vol. 80, no. 18, pp. 27187-27221, 2021.

[24] Malik, M., Malik, M. K., Mehmood, K. and Makhdoom, I., "Automatic speech recognition: a survey", *Multimedia Tools and Applications*, vol. 80, pp. 9411-9457, 2021.

[25] Li, J., "Recent advances in end-to-end automatic speech recognition", *APSIPA Transactions on Signal and Information Processing*, vol. 11, no. 1, 2022.

[26] Guo, Z., Leng, Y., Wu, Y., Zhao, S. and Tan, X, "PromptTTS: Controllable text-to-speech with text descriptions," *In ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1-5. IEEE, 2023.

[27] Saranya, V., Devi, T. and Deepa, N, "Text Normalization by Bi-LSTM Model with Enhanced Features to Improve Tribal English Knowledge", *In 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS),* pp. 1674-1679, 2023.

[28] Ultralytics YOLOv8, 2024. 01. 01. https://docs.ultralytics.com/.

[29] CoCo Dataset, 2024. 01. 01. https://cocodataset.org/.

[30] Google Cloud Speech, 2024. 01. 01. https://cloud.google.com/.