IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 1, December 2025
Impact Factor: 7.67

A Survey Paper on Human-Assisted Video Summary via Task Composition

Shreyas V, Uday Kumar J V, Thippesha T, Abhijeeth G, Shivraj Veerappa Banakar
Department of Information Science & Engineering
Global Academy of Technology, Bengaluru, India
shreshreyuas@gmail.com, tthippeshhyr@gmail.com
udayudayjv@gmail.com, abhivinu012@gmail.com

Abstract: *The exponential growth of multimedia data across digital platforms has sparked an ever-increasing need for intelligent, automated video summarization systems that are capable of generating concise, emotionally engaging, and contextually relevant summaries. State-of-the-art practices for creating trailers and editing videos still rely on highly manual approaches, wherein editors go through hours of footage to identify significant scenes. This process is very time-consuming, labor-intensive, and biased by human judgment. It is definitely impractical for use on large-scale or real-time applications. This paper provides an extensive survey and in-depth analysis of human-in-the-loop, AI-assisted video summarization frameworks with a focus on emotion-based scene extraction and collaborative editing.*
The paper proposes a combined scheme: MTCNN for face detection, FaceNet for identity recognition, and CNNs for emotion classification. These deep learning models detect, track, and analyze emotional expressions throughout the frames to identify scenes with the most narrative and affectively important content. Further, the frame level processing and trailer compilation are done using OpenCV, while a Flask-based interactive interface is used by human editors to review and refine the AI-generated summaries in order to balance automation with creative input. This survey brings together thirteen key research works that cut across predictive modeling, multimodal emotion recognition, and collaboration between AI and humans. A clear demonstration of how human intuition, coupled with machine precision, can improve efficiency by reducing editing time as high as 70%, without sacrificing quality or emotional depth, is depicted in the results. It also establishes the fact that emotion-aware hybrid systems will eventually turn traditional video editing into an adaptive, scalable, intelligent process and open up a whole new dimension toward next- generation media production frameworks which can present emotionally resonant and narratively cohesive video summaries.

Keywords: *multimedia data*

Copyright to IJARSCT
www.ijarsct.co.in

DOI: 10.48175/IJARSCT-30156