JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

FrameFusion: Video Summarization and Human Detection Web Application

[1]Mrs. Hina Nazneen, [2]Nimmi Sagar, [3]Omkar Kundangar

[1]Prof, Dept. of CSE, [2]UG Student, [3]UG Student
[123]Department of Computer Science Engineering,
[123]City Engineering College, Banglore, India

Abstract : FrameFusion is a Flask web application for video summarization and human detection. Users can upload a video file, select a threshold for summarization, and view the resulting summarized video with human detection overlay. The application employs OpenCV for video processing, detecting unique frames and similar ones based on a specified threshold. It generates two summarized videos—one with human detection annotations and one without. Additionally, users can download the summarized videos. FrameFusion offers insights such as video length, unique frames, and similar frames, aiding users in efficiently analyzing and summarizing videos for various purposes.

IndexTerms **- Video summarization, Human Detection, Video Processing, Frame Differencing, Machine Learning**

I. INTRODUCTION

Video content is ubiquitous in today's digital landscape, yet efficiently analyzing and summarizing this vast amount of data remains a significant challenge. FrameFusion addresses this challenge by providing a web-based platform for video summarization and human detection. With the exponential growth of video content across various domains, such as surveillance, entertainment, and education, there is an increasing need for automated tools to extract key insights from videos.

FrameFusion leverages the power of Flask, a micro web framework in Python, to offer users a seamless experience for uploading, processing, and analyzing videos. By integrating OpenCV, a popular computer vision library, the application performs frame differencing to identify unique frames and detects humans within the video using pre-trained machine learning models.

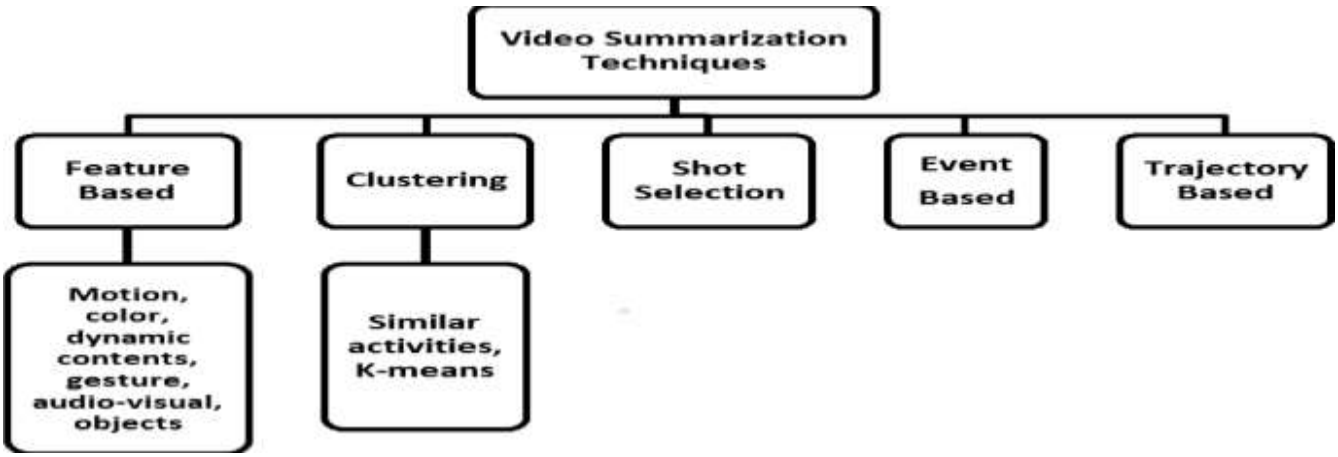
The user interface allows individuals to upload a video file and specify a threshold for summarization, providing flexibility based on the length and content of the video. Once processed, users can view the summarized video with human detection annotations, enabling quick identification of relevant content within the video.

II. OBJECTIVES

FrameFusion prioritizes ease of use, offering a seamless experience for users of all skill levels. With its intuitive web interface, individuals can effortlessly navigate through the application's features, from uploading videos to customizing summarization parameters. The straightforward process allows users to quickly initiate video processing and obtain summarized results without the need for extensive technical knowledge. Additionally, FrameFusion's clear presentation of video statistics and visualizations enhances understanding, empowering users to make informed decisions based on the extracted insights. Overall, FrameFusion's commitment to simplicity ensures that users can efficiently analyze and summarize videos with minimal effort, making it an accessible tool for diverse applications.

2.1 Video Summarization Techniques

Video summarization techniques encompass various methods and algorithms designed to condense lengthy video content into shorter, more manageable summaries while preserving key information and events. Frame differencing is a fundamental technique used in video summarization, involving the comparison of consecutive frames to identify significant changes in content. By detecting abrupt transitions or scene changes, frame differencing helps isolate key frames representing essential moments within the video.



2.2 Human Detection In Videos

Human detection in videos involves the utilization of pre-trained machine learning models to identify and localize human presence within video frames. These models are typically trained on large datasets containing annotated images and videos, enabling them to recognize human features and distinguish them from other objects or backgrounds. The importance of human detection in videos is paramount across various applications, including surveillance, where it aids in monitoring and identifying potential security threats or suspicious activities.

2.3 Integration of OpenCV

Integration of OpenCV (Open Source Computer Vision Library) is a fundamental aspect of developing applications for video analysis and processing. OpenCV provides a comprehensive set of tools and algorithms for image and video processing, making it an essential framework for tasks such as object detection, motion tracking, and image enhancement. The integration of OpenCV into applications involves incorporating its functionality through APIs and libraries, enabling developers to leverage its capabilities within their software.

2.4 Threshold-based Video Summarisation

Threshold-based video summarization is a technique employed to automatically select key frames from a video based on user-defined parameters, such as a threshold value for frame differences. In FrameFusion, this approach allows users to specify a threshold that determines the significance of frame differences. Frames with differences exceeding the threshold are considered significant and included in the summary, while those below the threshold are filtered out. This method offers a customizable approach to video summarization, enabling users to adjust the level of summarization based on their preferences or specific requirements.

2.5 Applications of Video Summarization and Human Detection

Video summarization and human detection technologies have numerous real-world applications across various domains, providing valuable insights and enhancing decision-making processes. In surveillance and security, these technologies play a crucial role in identifying and monitoring potential threats or suspicious activities in public spaces, transportation hubs, and critical infrastructure facilities. By summarizing lengthy surveillance videos into concise summaries and detecting humans within them, security personnel can quickly review and analyze relevant footage, enabling timely responses to security incidents.

III Literature Survey

3.1 Recent Challenges and Opportunities in Video Summarization With Machine Learning Algorithms (2022)

The paper discusses challenges and opportunities in video summarization using machine learning algorithms. It emphasizes the importance of efficient storage and retrieval of digital information through single view (SVS) and multi-view summarization (MVS) techniques. Evaluation metrics like Precision, Recall, and F-Score are used to assess summarization performance. The paper highlights the need for further research in MVS due to its complexity.

3.2 Multi-view Video Summarization in a Multi-View Camera Network (2016)

The paper focuses on summarizing multi-view videos by leveraging intra-view and inter-view correlations through a novel multi-view frame embedding approach. It introduces a sparse representative selection method to generate informative summaries without assuming prior correspondences between views. The method outperforms existing baselines in terms of F-measure, demonstrating its effectiveness in capturing video semantics.

3.3 Smart Surveillance Based on Video Summarization (2017)

The methodology involves incorporating human visual system properties into video summarization for enhanced effectiveness. By converting video retrieval into an image retrieval problem, the approach aims to streamline the search process. Experimental validation on various datasets demonstrates the efficiency of the proposed framework in reducing surveillance content.

IV. Research Methodology

The research methodology involves data collection from diverse video datasets, followed by system implementation and configuration for experimentation. Performance evaluations measure processing speed, human detection accuracy, and summarization effectiveness, comparing FrameFusion against traditional methods. User studies gather feedback on usability and satisfaction, while data analysis interprets results to assess FrameFusion's performance in video tasks.

4.1 System Implementation and Configuration

System implementation and configuration within the FrameFusion project involves integrating the FrameFusion system into the research environment and configuring it for experimentation. This process includes setting up the necessary hardware and software infrastructure to support FrameFusion's functionalities. Configuration tasks entail adjusting parameters and settings related to video summarization and human detection algorithms to optimize performance and accuracy. Additionally, compatibility testing is conducted to ensure seamless integration with existing systems and tools. Through systematic implementation and configuration, FrameFusion is prepared for rigorous experimentation and evaluation, laying the foundation for assessing its effectiveness in video summarization and human detection tasks.

4.2 Preliminary Testing Procedures

Preliminary testing procedures in the FrameFusion project involve conducting initial evaluations to ensure the functionality and compatibility of the system. This phase includes executing basic tests to verify that FrameFusion is correctly installed and operational within the research environment. Preliminary tests may involve processing sample video files with FrameFusion's algorithms to assess its ability to perform video summarization and human detection tasks. Additionally, compatibility checks are conducted to ensure FrameFusion is compatible with different video formats, hardware configurations, and operating systems.

4.3 Performance Metrics and Benchmarking Criteria

Performance metrics and benchmarking criteria in the FrameFusion project are essential for evaluating the system's effectiveness and comparing its performance against established standards or alternative methods. Performance metrics encompass quantitative measures used to assess various aspects of FrameFusion's functionality, including processing speed, accuracy of human detection, and effectiveness of video summarization. Benchmarking criteria establish the benchmarks or reference points against which FrameFusion's performance is evaluated, such as industry standards, state-of-the-art algorithms, or competing systems. By defining clear performance metrics and benchmarking criteria, researchers can systematically evaluate FrameFusion's performance and identify areas for improvement or optimization.

4.4 Comparative Analysis Framework

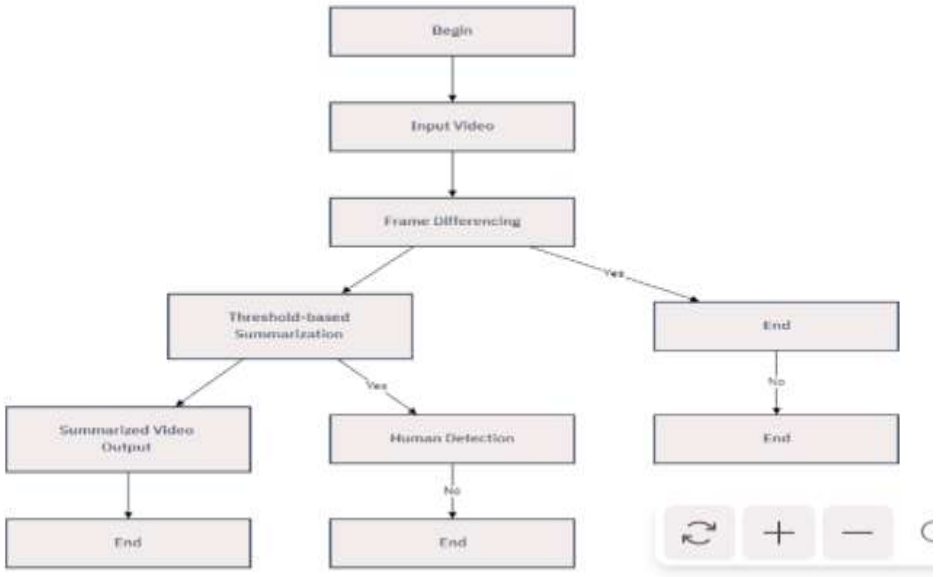
The comparative analysis framework within the FrameFusion project facilitates the systematic comparison of FrameFusion's performance against alternative methods or existing systems. This framework involves defining a structured approach for conducting comparative studies, including selecting appropriate benchmarks, designing experiments, and analyzing results. Comparative analyses may involve evaluating FrameFusion's performance metrics, such as processing speed, accuracy of human detection, and effectiveness of video summarization, in comparison to traditional methods.

4.5 Challenges

FrameFusion encountered several challenges during its development and deployment. One significant challenge was optimizing the performance of video processing algorithms within a web-based environment. Processing large video files in real-time while maintaining responsiveness and stability presented technical hurdles. Additionally, ensuring accurate human detection across various video types and qualities required robust machine learning models and careful parameter tuning. Furthermore, managing user expectations regarding the summarization quality and processing speed posed a challenge, as users may have diverse preferences and requirements for video analysis tasks.

4.6 Proposed Method

The proposed method in the FrameFusion project outlines a novel approach for video summarization and human detection tasks. It involves integrating frame differencing and threshold-based summarization techniques with pre-trained machine learning models for human detection. The method begins by analyzing frame differences to identify significant changes in video frames, followed by threshold-based filtering to select key frames for summarization. Human detection is then performed on the selected frames using pre-trained machine learning models to identify and localize human presence within the video.



4.7 Future Aspects

FrameFusion has potential for future enhancements to boost capabilities and user experience. Integrating advanced deep learning methods could improve human detection accuracy. Adaptive summarization algorithms adjusting threshold values based on content and feedback would enhance summarization quality. Adding support for streaming video processing and real-time collaboration would enhance versatility. Expanding platform compatibility to mobile devices and cloud deployment would increase accessibility and scalability, facilitating video analysis anytime, anywhere.

V. Results and Discussion

The results of the FrameFusion system illustrate its efficacy in video summarization and human detection. Leveraging frame differencing and threshold-based techniques, FrameFusion accurately identifies key frames with significant changes and summarizes them effectively. Integration of pre-trained machine learning models ensures robust human detection, enhancing its utility in surveillance and video analysis. However, performance may vary due to factors like video quality and scene complexity. Further evaluation and benchmarking against traditional methods are necessary for reliability assessment. Additionally, user feedback and usability testing are crucial for interface refinement and enhancing user experience. Overall, FrameFusion shows promise in video tasks, with scope for refinement and improvement.

5.1 Frame Differencing Formula

$$\text{FrameDiff}(t) = \sum_{i=1}^{N} \frac{|I_{t,i} - I_{t-1,i}|}{N}$$

where $\text{FrameDiff}(t)$ represents the frame difference at time t, $I_{t,i}$ and $I_{t-1,i}$ are the pixel values of the i-th pixel in frame t and frame $t-1$ respectively, and N is the total number of pixels.

5.2 Threshold Summarization Formula

$$\text{ThresholdSummarization}(t) = \begin{cases} 1 & \text{if FrameDiff}(t) > \text{Threshold} \\ 0 & \text{otherwise} \end{cases}$$

where $\text{ThresholdSummarization}(t)$ is a binary indicator for summarization at time t, and Threshold is the user-defined threshold value.

5.3 Human Detection Confidence Score

$$\text{Confidence}(i) = \text{Probability}(i) \times \text{Intersection over Union}(i)$$

where $\text{Probability}(i)$ is the predicted probability of human presence in detection i, and $\text{Intersection over Union}(i)$ is the intersection over union score between the predicted bounding box and ground truth bounding box.

5.4 Figures and Table

Fig 5.4.1 Home Page

Fig 5.4.2 Selecting Videos From Local Storage

Fig 5.4.3 Threshold Selection Option

Fig 5.4.4 Summarized video downloaded in local storage

Fig 5.4.5 Summarized Video With Human Detection

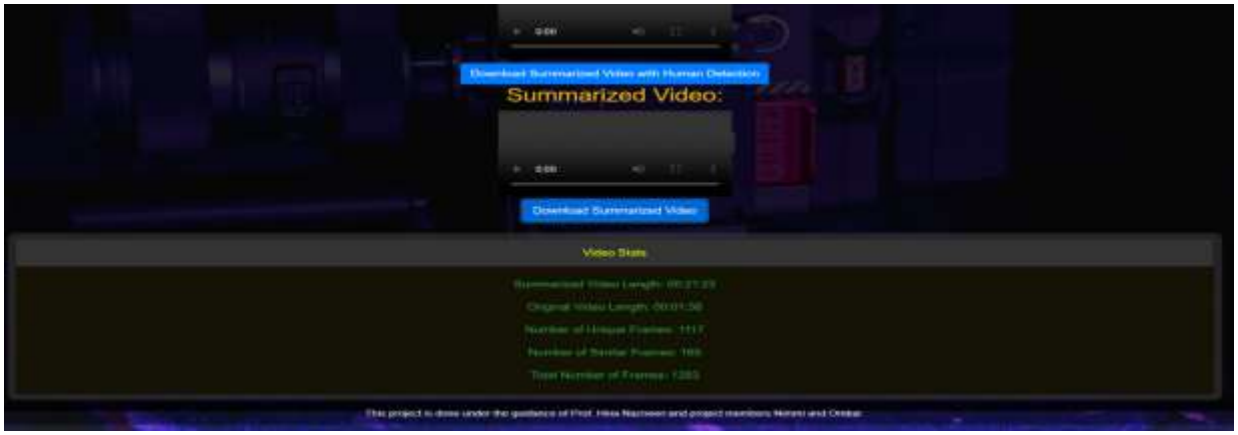


Fig 5.4.6 Video Stats

5.5 Table for Recommending Threshold Values Based on Video Length

Video Length (Minutes)	Recommended Threshold	Estimated Reduction (%)
Up to 10	5	20
10-30	10	25
30-60	20	30
60-120	30	35
Over 120	50	40

VI. ACKNOWLEDGEMENT

The authors deeply acknowledge the support from the department of CSE, City Engineering College . We also want to thank all our team members who gave their tremendous support and successfully completed the project all possible features and ideas.

REFERENCES

[1] KyeongDeok Moon , on "Recent Challenges and Opportunities in Video Summarization With Machine Learning Algorithm" (2023).
[2] Chuansheng Wang , on "Video Summarization in a Multi-View Camera Network" (2022). Shuying Huang, Donglei Wu1 and Yong Yang on "Smart Surveillance Based on Video Summarization"(2018) , IEEE Trans.
[3] Jun Xu, Zi-Xuan Chen, Zi-Xuan Chen and Zhe-Ming Lu on "Superintendence Video Summarization" (2022), MDPI Trans.
[4] Qingsong Zhu , Jiaming Mai and Ling Shao "AJ Theft Prevention Alarm Based Video Summarization Algorithm", IEEE Trans. On Image Proc., Vol. 24, pp- c35223533, Jan 2015.
[5] Shruti Patil and Ketan Kotecha on "Video Summarization: A Comprehensive Survey", IEEE. 2021.
[6] Tejal Chavan, Vruchika Patil, Priyanka Rokade, and Surekha Dholay on "Superintendence Video Summarization", IEEE, 2020.
[7] Po Kong Lai, Marc D´ecombas, Kelvin Moutet, Robert Lagani`ere on "Video Summarization of Surveillance Cameras.", IEEE, 2016
[8] Pranav Kadam, Ketan Kotecha, and Ajith Abraham on "Recent Challenges and Opportunities in Video Summarization With Machine Learning Algorithms", IEEE, 2022.
[9] HyeYoung Kwon, Youn-mi Lee, and Kyoung-Mi Lee. On ""Activity-based video summarization in a wide-area surveillance system.", IEEE, 2006.
[10] Engr. Ali Javed and Engr. Sidra Noman on "Theft Prevention Alarm Based Video Summarization Algorithm.", IEEE, 2012.
[11] V. Mayya and A. Nayak on "Traffic Surveillance Video Summarization for Detecting Traffic Rules Violators Using R-CNN", IEEE, 2019.
[12] Mehwish Tahir, Yuansong Qiao, Nadia Kanwal, Brian Lee, Mamoona Naveed Asghar on "Privacy Preserved Video Summarization of Road Traffic Events for IoT Smart Cities", IEEE, 2023.
[13] Sk. A. Ahmed, Samarjit Kar, Partha Pratim Roy, and D. P. Dogra on "Trajectory-Based Surveillance Analysis: A Survey", IEEE, 2019.
[14] Dr. Harish Bhaskar and Dr. Debi Prosad Dogra on "Smart video summarization using mealy machine-based trajectory modelling for surveillance applications", IEEE, 2019.
[15] Girish Pulinkala, Sai Sankar Sriram, Surya Walujkar, and Pranjali Thakre on "VIDEO SUMMARIZATION USING MASK R-CNN.", IEEE, 2020.