Video Processing Approaches: A Review

Mayur Akewar, *Alumni* [1]
Department of Computer Science & Engineering
Shri Ramdeobaba College of Engineering & Management
Email: [1]*mayurakewar87@gmail.com*

***Abstract*—Video processing is a fundamental task in computer vision and multimedia analysis, encompassing various techniques for capturing, analyzing, and manipulating video data. In this paper, we conduct a comprehensive review of video processing approaches, covering topics such as video enhancement and video content detection. We examine the evolution of video processing techniques, discuss their applications across different domains, and highlight current challenges and future research directions.**

***Index Terms*—Video Processing**

I. Introduction

Video processing plays a crucial role in extracting meaningful information from video data, enabling tasks such as surveillance, entertainment, healthcare, and education. Over the years, significant advancements have been made in video processing techniques, driven by developments in computer vision, machine learning, and multimedia technologies. In this paper, we present a review of video processing approaches, focusing on the evolution of techniques, their applications in various domains, and the challenges faced by researchers in the field. By synthesizing existing literature and analyzing recent advancements, we aim to provide insights into the state-of-the-art techniques and future directions in video processing research.

II. Literature Review

The foundation of our work in video analysis and intelligence rests upon a broad body of literature that explores the intersection of multi-modal data analysis and video content understanding. In this section, we delve into key studies and methodologies that inform our proposed framework. Our review encompasses pivotal research in text, image, and audio analysis as applied to video data, shedding light on the evolution of multi-modal approaches in the field. In [5], the authors present the USTC_SmokeRS benchmark for smoke detection in satellite imagery using MODIS data. The dataset contains 6225 images across six classes and is used to evaluate deep learning models. A novel CNN model, SmokeNet, is introduced, which incorporates spatial and channel-wise attention. SmokeNet outperforms other methods in accuracy and Kappa coefficient, with the best results, achieved using 64% of the training images. In [6], the authors present a novel training method, Bag of Focus (BoF), for deep neural networks in surveillance video analysis. BoF involves training on motion-intensive video blocks, reducing computational costs by 90% without sacrificing performance. It includes the creation of a detailed dataset for real-world volume crimes. The study shows that even 2D Convolutional Neural Networks (CNNs) with fewer parameters achieve a high classification accuracy of 98.7% and an Area under the Curve (AUC) of 99.7% for recognizing malicious events in videos, outperforming some 3D CNNs. In [7], the authors explores on-camera filtering for real-time video analytics, addressing resource constraints. It introduces a system called Reducto, which adapts filtering decisions based on feature type, threshold, query accuracy, and video content. Reducto significantly reduces frame processing (51-97%) while maintaining accuracy across different videos and queries, mitigating resource demands. In [8], the authors address the challenges of handling vast surveillance camera data, particularly in prison settings, and the importance of intelligent video analytics. It introduces a method that utilizes deep learning, specifically the Mask R-CNN framework, for object detection and segmentation in video images. Experimental results showcase the method's effectiveness, achieving a high mask average precision of nearly 98.5% on their datasets.

In [9], the authors focus on analyzing video streams from intelligent transportation systems and introduce a long video event retrieval algorithm based on superframe segmentation. This method effectively reduces redundancy by detecting motion amplitude, segments video into Segments of Interest (SOIs) with feature fusion, and employs a semantic model for text-to-video matching. Experiments demonstrate improved efficiency, accuracy in semantic description, and reduced retrieval time for these transportation videos. In [10], the authors present a thorough survey on the use of deep learning in surveillance video analysis, with a focus on object recognition, action recognition, crowd analysis, and violence detection in crowded settings. It highlights challenges in counting, identifying individuals, and recognizing activities in large crowds under different weather conditions. The survey covers deep learning-based methods, discusses their algorithms and models, and also points out current issues while suggesting directions for future research to tackle these challenges. In [11], the authors present an end-to-end framework for autonomous vehicles (AVs) to follow other vehicles. It uses RGB-D frames, object detection with YOLOv3, and reinforcement learning (RL) algorithms (Q-learning and Deep Q-learning) for navigation. Simulation results show that AVs can effectively follow cars using only video frames, demonstrating reasonable car-following behavior. In [12], the authors introduce a novel method for monitoring the elderly, focusing

on the detection of abnormal behaviors, unusual events, and daily activities. This approach combines deep convolutional features with joint Bayesian modeling to account for different perspectives. It is tested on two behavioral datasets and proves to be more effective than existing methods, demonstrating its suitability for elderly care monitoring. In [13], the authors introduce an event summarization method for monocular videos, specifically in video surveillance. It employs deep learning and a spatiotemporal similarity function to create a similarity matrix based on visual features. The approach forms Highly Connected Subgraphs (HCS) as clusters, and events are determined from these clusters using cluster centroids as keyframes. Notably, this method doesn't require pre-specifying the number of clusters, offering flexibility. Experimental results on benchmark datasets show that this model excels in terms of Precision and F-measure while effectively summarizing the original video content. In [14], the authors present a novel approach called Spatio-Temporal Completion network (STCnet) to address partial occlusion in video person re-identification in surveillance videos. STCnet can recover the appearance of occluded body parts by utilizing spatial and temporal information, improving re-ID accuracy by combining the recovered parts with unoccluded ones. This approach, referred to as Video Re-ID Framework Robust to Partial Occlusion (VRSTC), surpasses existing methods on challenging video re-ID databases. In [15], the authors focus on Human Action Recognition (HAR), important for real-time applications like video surveillance and rescue missions. HAR faces challenges due to variations in human attributes, lighting, and camera settings. The study introduces HAREDNet, a hybrid recognition approach that excels in feature extraction and achieves high recognition accuracy. It outperforms previous methods with impressive scores on three datasets: 97.45%, 80.58%, and 97.48% on NTU RGB+D, HMDB51, and UCF-101, respectively. In [16], the authors tackle limitations in video database management systems (VDBMSs) that use deep learning for video analysis. It presents FiGO, which overcomes these limitations through an ensemble of models, precise query optimization, and lightweight pruning. FiGO outperforms existing systems, delivering a notable 3.3x improvement in query processing efficiency across four video datasets on average. In [17], the authors proposed a video analytics system Vaas for large-scale datasets that provides an interactive platform for creating and testing video analytics workflows. Users can define these workflows as Vaas queries, which encompass machine learning models and various operations. The system offers tools for query composition and experimentation with dataset samples. It uses approximate video query processing for speed and interactivity, and it streamlines the annotation process by enabling users to annotate over the outputs of previous queries instead of the entire video dataset.

III. Conclusion

Video processing approaches play a vital role in extracting valuable information from video data, enabling a wide range of applications across different domains. Through our review, we have explored the evolution of video processing techniques, from traditional methods to more advanced approaches based on deep learning and artificial intelligence. We have discussed the applications of video processing in areas such as surveillance, entertainment, healthcare, and education, highlighting the diverse range of tasks that can be accomplished through video analysis. Despite significant advancements, challenges such as robustness to variations in lighting and scene complexity, real-time processing requirements, and privacy concerns remain areas of ongoing research. Moving forward, addressing these challenges and continuing to innovate in video processing will pave the way for new opportunities and advancements in multimedia analysis and computer vision.

References

[1] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, Ilya Sutskever *Learning Transferable Visual Models From Natural Language Supervision* . arXiv, 2021. [Online]. Available: https://arxiv.org/abs/2103.00020

[2] M. Plakal and D. Ellis *Yamnet* , 2020. [Online]. Available: https://github.com/tensorflow/models/tree/master/research/audioset/yamnet

[3] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, Ilya Sutskever *Robust Speech Recognition via Large-Scale Weak Supervision* . arXiv, 2022, [Online]. Available: https://arxiv.org/abs/2212.04356

[4] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, Luke Zettlemoyer *BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension* . arXiv, 2019. [Online]. Available: https://arxiv.org/abs/1910.13461

[5] Rui Ba, Chen Chen, Jing Yuan, Weiguo Song, Siuming Lo *SmokeNet: Satellite Smoke Scene Detection Using Convolutional Neural Network with Spatial and Channel-Wise Attention* . Remote Sensing, 2019. [Online]. Available: https://www.mdpi.com/2072-4292/11/14/1702

[6] Atif Jan, Gul Muhammad Khan *Real World Anomalous Scene Detection and Classification using Multilayer Deep Neural Networks* . International Journal of Interactive Multimedia and Artificial Intelligence (IJIMAI), 2023. [Online]. Available: https://reunir.unir.net/handle/123456789/14335

[7] Yuanqi Li, Arthi Padmanabhan, Pengzhan Zhao, Yufei Wang, Guoqing Harry Xu, Ravi Netravali *Reducto: On-Camera Filtering for Resource-Efficient Real-Time Video Analytics* . Proceedings of the Annual conference of the ACM Special Interest Group on Data Communication on the applications, technologies, architectures, and protocols for computer communication, 2020. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3387514.3405874

[8] Yun-Xia Liu, Yang Yang, Aijun Shi, Peng Jigang, Liu Haowei *Intelligent monitoring of indoor surveillance video based on deep learning* . 21st International Conference on Advanced Communication Technology (ICACT), 2019. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8701964

[9] Shaohua Wan, Xiaolong Xu, Tian Wang, Zonghua Gu *An Intelligent Video Analysis Method for Abnormal Event Detection in Intelligent Transportation Systems* . IEEE Transactions on Intelligent Transportation Systems, 2021. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9190063

[10] G. Sreenu, M. A. Saleem Durai *Intelligent video surveillance: a review through deep learning techniques for crowd analysis* . Springer, Journal of Big Data, 2019. [Online]. Available: https://link.springer.com/article/10.1186/s40537-019-0212-5

[11] Mehdi Masmoudi, Hamdi Friji, Hakim Ghazzai, Yehia Massoud *A Reinforcement Learning Framework for Video Frame-Based Autonomous Car-Following* . IEEE Open Journal of Intelligent Transportation Systems, 2021. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9439506

[12] Xiujun Gao *Abnormal Behavior Detection and Warning Based on Deep Intelligent Video Analysis for Geriatric Patients* . Journal of Medical Imaging and Health Informatics, 2021.

[13] Krishan Kumar *EVS-DK: Event video skimming using deep keyframe* . Journal of Visual Communication and Image Representation, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S1047320318303353

[14] Ruibing Hou, Bingpeng Ma, Hong Chang, Xinqian Gu, Shiguang Shan, Xilin Chen *VRSTC: Occlusion-Free Video Person Re-Identification* . Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[15] Inzamam Mashood Nasir, Mudassar Raza, Jamal Hussain Shah, Shuihua Wang, Usman Tariq, Muhammad Attique Khan *HAREDNet: A deep learning based architecture for autonomous video surveillance by recognizing human actions* . Computers and Electrical Engineering, 2022 [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0045790622001057

[16] Jiashen Cao, Karan Sarkar, Ramyad Hadidi, Joy Arulraj, Hyesoon Kim *FiGO: Fine-Grained Query Optimization in Video Analytics* . Proceedings of the 2022,International Conference on Management of Data 2022. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3514221.3517857

[17] Favyen Bastani, Oscar Moll, Sam Madden *Vaas: Video Analytics At Scale* . Proceedings of the VLDB Endowment, 2020.

[18] Zhiquan Ye, Yuxia Geng, Jiaoyan Chen, Jingmin Chen, Xiaoxiao Xu, SuHang Zheng, Feng Wang, Jun Zhang, Huajun Chen *Zero-shot Text Classification via Reinforced Self-training* . Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020. [Online]. Available: https://aclanthology.org/2020.acl-main.272/

[19] Robert Gorwa , Reuben Binns, Christian Katzenbach *Algorithmic content moderation: Technical and political challenges in the automation of platform governance*. Big Data & Society, 2020.

[20] Learning Transferable Visual Models From Natural Language Supervision.[Online]. Available: https://github.com/openai/CLIP

[21] Robust Speech Recognition via Large-Scale Weak Supervision.[Online]. Available: https://github.com/openai/whisper

[22] BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. [Online]. Available: https://huggingface.co/facebook/bart-large-mnli