AI-Based Video Analysis and Automated Highlight Generation for Bike Ride Videos

Dr.V.Deepa[1], Mr.AghashKumar[2], Mr.S.Keerthivasan[3]
[1]*Professor, Department of Computer Science, Hindusthan College of Arts & Science, Coimbatore,641028*
[2,3]*IIPG Student, Department of Computer Science, Hindusthan College of Arts & Science, Coimbatore,641028*

Abstract: The growing popularity of bike ride and moto-vlog videos has increased the need for efficient video editing solutions. Long-duration ride footage often contains only a few engaging moments, making manual highlight extraction time-consuming. This project presents an AI-based video analysis system that automatically identifies and extracts important segments from bike ride videos. The proposed system combines computer vision techniques and audio signal processing to analyze scene transitions, motion intensity, and engine sound levels. A scoring mechanism is used to evaluate each segment, and high-scoring portions are selected to generate a summarized highlight video. The system reduces manual editing effort while preserving engaging content, making it useful for travel vloggers and digital content creators.

Keywords: Video Analysis, Highlight Detection, Motion Estimation, Scene Detection, Audio Processing, Artificial Intelligence

I. INTRODUCTION

Digital platforms have significantly increased the creation and sharing of bike ride videos. Thesevideosareoftenrecordedcontinuously for long durations, but only specific segments contain visually interesting or dynamic content. Editing such videos manually requires reviewing the entire footage, which is time-intensive and inefficient.

Advancements in Artificial Intelligence and computer vision have enabled automated multimedia analysis. By analyzing visual motion, scene changes, and audio intensity, it is possible to identify important events within a video. This project focuses on designing a lightweight automated system that extracts meaningful highlights from bike ride videos using visual and audio feature analysis. The goal is to reduce editing time while maintaining content quality.

II. LITERATURE SURVEY

1. Study on Video Summarization Techniques: Previous research on video summarization focuses on generating short and meaningful representations of long- duration videos. Traditional approaches rely on key-frame extraction, frame differencing, and shot boundary detection to identify important segments. These techniques mainly analyze visual variations between consecutive frames. While effective for basic summarization, they may fail to capture semantic importance in dynamic videos.

2. Research on Motion and Scene Analysis in Video Processing: Several studies have explored motion estimation and scene transition detection for activity recognition. Optical flow and histogram-based comparison methods are widely used to measure movement intensity and detect scene changes. These techniques are computationally efficient and suitable for real-time applications. However, they often depend heavily on threshold selection and may require tuning based on video type.

3. Audio-Based Event Detection in Multimedia Systems: Recent research highlights the importance of audio signal analysis in video understanding. Techniques such as amplitude envelope detection, short- term energy computation, and frequency spectrum analysis are used to detect significant sound events. Audio-based methods enhance highlight detection accuracy when combined with visual analysis. However, background noise and environmental sounds can affect performance if not properly filtered.

III. PROBLEM STATEMENT

Long-duration bike ride videos contain limited segments of high engagement. Manual identification of these segments is inefficient and time-consuming. There is a need for an automated system that can analyze both visual and audio components of a video to detect important ride moments and generate concise highlight clips.

IV. METHODOLOGY

The proposed system follows a multi-stage processing architecture consisting of input handling, feature extraction, scoring, and highlight generation.

1. Video Input and Pre-Processing

The system accepts a raw bike ride video file as input. The video is divided into smaller segments for analysis. Two pre-processing operations are performed:

- Frame extraction at fixed intervals
- Audio signal separation from the video file

This separation enables independent analysis of visual and audio components.

2. Visual Feature Analysis

a) Scene Transition Detection

Consecutive frames are compared using histogram difference techniques. Significant variation between frames indicates scene changes such as environmental shifts or camera angle changes.

b) Motion Intensity Estimation

Frame differencing and optical flow estimation are used to compute motion magnitude. High motion intensity typically corresponds to acceleration, sharp turns, or overtaking maneuvers.

c) Brightness and Contrast Evaluation

Changes in illumination levels are analyzed to detect visually appealing scenic moments such as sunrise or sunset rides.

3. Audio Feature Extraction

The extracted audio waveform is analyzed to measure:

- Signal amplitude
- Short-term energy
- Frequency distribution

High engine acceleration and dynamic riding conditions often produce increased audio intensity, which contributes to highlight scoring.

4. Feature Fusion and Scoring Mechanism

The system integrates visual and audio features using a weighted scoring formula:

Highlight Score = (w1 × Motion Score) +(w2 × Scene Change Score) + (w3 × Audio Energy Score)

Where w1, w2, and w3 represent predefined weight values.

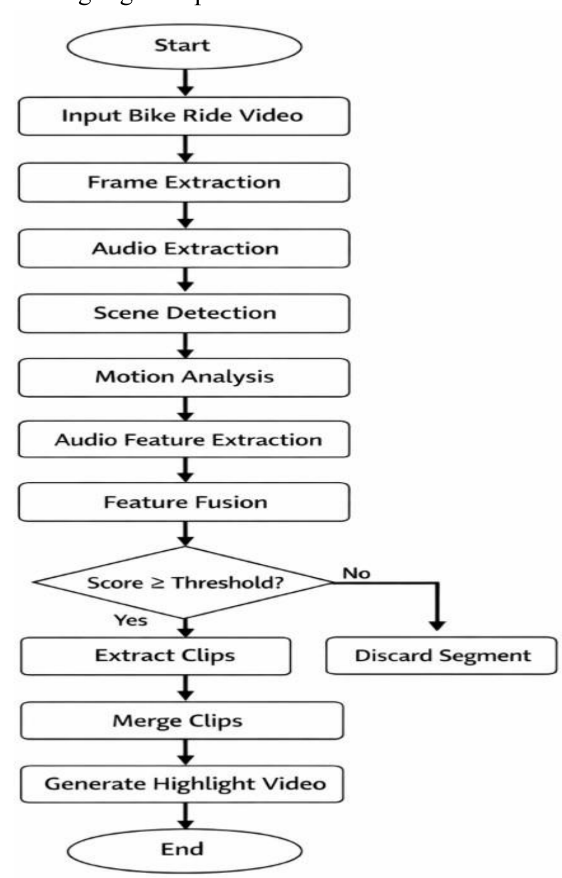
If the computed highlight score exceeds a specified threshold, the segment is classified as important.

5. Highlight Generation

Selected segments are extracted from the original video timeline. These clips are sequentially merged to generate a summarized highlight video. The final output retains only the most engaging portions of the ride.

V. FLOWCHART

The flowchart illustrates the sequential execution of the automated highlight generation process. It represents the logical workflow from video input to final highlight output.



VI. EXPERIMENTAL SETUP

Data Collection

Real-world bike ride videos recorded in urban and highway environments were used for testing. The videos included varying lighting conditions and riding speeds.

Tools and Technologies Used

- Python for system implementation
- OpenCV for computer vision operations
- Librosa for audio signal processing
- MoviePy for video editing and clip merging
- NumPy for numerical computations

Evaluation Metrics

- Processing time
- Duration reduction ratio
- Accuracy of highlight detection (based on manual validation)
- User feedback satisfaction

VII. RESULTS AND DISCUSSION

The proposed system successfully identified dynamic riding segments and scenic views from long-duration videos. On average, the generated highlight video was 60–70% shorter than the original footage.

Motion-based detection accurately captured high-speed and turning events. Audio energy analysis effectively identified engine acceleration moments. Scene detection contributed to detecting environmental transitions.

However, certain limitations were observed:

- Extremely low-light conditions reduced motion detection accuracy.
- Background noise occasionally affected audio-based scoring.
- Threshold tuning required adjustment for different riding styles.

Despite these limitations, the system significantly reduced manual editing effort while maintaining engaging content quality.

VIII. CONCLUSION

This research introduced an AI-based video analysis framework designed to automatically generate highlights from bike ride videos. By combining computer vision techniques and audio signal processing methods, the system effectively identifies and extracts important segments without requiring complex deep learning model training.

The modular architecture and lightweight implementation make the system suitable for practical use by moto-vloggers and travel content creators. The results demonstrate that automated highlight generation can enhance workflow efficiency and reduce editing time.

IX. FUTURESCOPE

Future enhancements may include:

- Integration of deep learning-based action recognition
- Real-time highlight detection during live recording
- Automatic background music synchronization
- Cloud-based deployment for large- scale content creators

REFERENCES

[1] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 4th ed. Pearson, 2018.

[2] A. K. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1989.

[3] B. K. P. Horn and B. G. Schunck, “Determining optical flow,” *Artificial Intelligence*, vol.17, no.1–3, pp.185–203, 1981.

[4] J. Canny, “A computational approach to edge detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[5] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *Proc. Int. Conf. Learning Representations (ICLR)*, 2015.

[6] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2014, pp. 2672–2680.
[7] D. G. Lowe, "Distinctive image features from scale-invariant key points," *International Journal of Computer Vision*, vol.60, no.2, pp.91–110, 2004.
[8] T. Giannakopoulos and A. Pikrakis, *Introduction to Audio Analysis: A MATLAB Approach*. Academic Press, 2014.
[9] A. Rosebrock, *Practical Python and OpenCV*. PyImage Search, 2016.
[10] R. Szeliski, *Computer Vision: Algorithms and Applications*, 2nd ed. Springer, 2022.