Overlap Detection via Audio Fingerprinting & Deep Learning

Technical Brief

Ref No: TechBrief/2025/18

Technology Summary

This innovative system automatically detects overlapping content between two videos by analyzing their audio tracks using audio fingerprinting and a lightweight deep learning model. It excels in identifying overlapping segments in unconstrained, multi-user, performance-event recordings—even under noisy conditions improving alignment accuracy by an average of 13.71% over existing methods.

Background

Synchronizing multiple video recordings from the same event such as concerts or performances is challenging, especially when done by different users with no shared timestamps, varying recording lengths, and background noise. Visual only approaches often fail due to differing camera angles. By focusing on audio overlap detection as a preprocessing step, this method provides a robust foundation for automated video alignment without manual intervention.

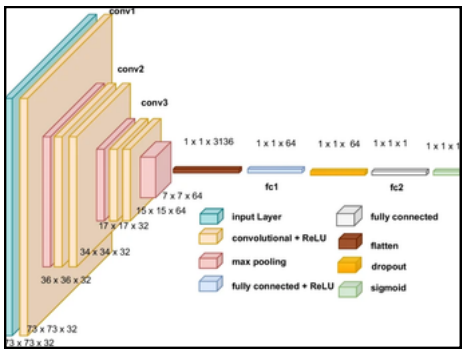
Technology Description

The method operates in two phases:

1. Audio Fingerprinting (Phase 1): Generate fingerprints by converting audio into spectrograms, extracting "constellation map" peak frequencies, and creating paired time-frequency hash keys (landmarks) stored in a hash table.
2. Deep Learning–Based Overlap Detection (Phase 2): Compare landmark sets from two audio clips to derive matching time-offset scatter plots. These plots are fed to a compact CNN (three convolutional layers + pooling + dense layers) that classifies whether segments overlap.
3. The convolutional model is lightweight and robust, suitable for deployment in devices with limited compute.

Market Potential / Proposed Deployment

- Market Trends: Demand growing for automated content synchronization tools in media editing, video summarization, and multimedia retrieval.
- Target Sectors: Event videography platforms, media production tools, surveillance systems, and social media aggregators.
- Socio-Economic Impact:
- Reduces editing labor costs by automating synchronization workflows.
- Enables seamless creation of high-quality mashups and coverage compilations from multiple user sources.



Applications

- Automated Video Stitching: Align and merge multiple user recordings of the same performance without manual tagging.
- Surveillance & Content Moderation: Detect overlapping footage in multi-source surveillance or user-generated content.
- Media Retrieval & Summarization: Aid in identifying redundant segments or common content across vast video archives.

Value Proposition

- Robust: High resilience to noise and variability in real-world audio captures.
- Efficient: Lightweight CNN ensures rapid classification on moderate hardware.
- Improved Accuracy: Outperforms standard audio-alignment methods by over 13%.
- Automated: Eliminates need for manual tagging and synchronization inputs.

Technology Status

- TRL Level: 4–5 foundational research validated on real-world dataset (Jiku Mobile Video Dataset).
- Outcome: Prototype system with proven superiority versus DTW and standard deep learning baselines.
- IP Status: Published in peer-reviewed journal; no patent recorded so far.