

# VIDEO INDEXING

STEP 2:

## FEATURE EXTRACTION





## Purpose of Feature Extraction



**Identify visual elements (objects, colors, movements, faces).**

**Extract audio content (spoken words, background noise, music, silence).**

**Detect on-screen text (subtitles, captions).**



Techniques Used



# DOMINANT COLOR EXTRACTION



## TOOL

OpenCV, scikit-learn's K-means.



## HOW

Clusters pixels to find dominant colors in keyframes.



## WHY

Describes scene appearance (e.g., "green park").



Techniques Used



# MOTION DETECTION



## TOOL

OpenCV (Lucas-Kanade optical flow).



## HOW

Tracks corner movements between frames to measure motion.



## WHY

Captures dynamic actions (e.g., “dog running”).



Techniques Used



## TOOL

NumPy, SciPy, Matplotlib.



## HOW

Computes RMS energy from audio to measure intensity.



## WHY

Identifies audio events (e.g., loud barks).





Techniques Used



# SPEECH-TO- TEXT



**TOOL**

SpeechRecognition  
(PocketSphinx).



**HOW**

Converts audio to text using offline  
speech recognition.



**WHY**

Extracts spoken words for indexing.



## Importance of Step 2



**DRIVES INDEXING**



**SUPPORTS MULTIMODAL PROCESSING**



**ENABLES SEGMENTATION**



**PREPARES FOR METADATA**



- Features like “dog” or “park” become keywords for search.
- Example: A frame with a dog and subtitle “park” is tagged for both.

Combines visual (histograms) and audio (energy) cues for robust keyframe detection

Features help divide videos into meaningful segments (Step 3).

Raw features feed into metadata consolidation (Step 4).

