

VIDEO INDEXING

STEP 2:

FEATURE EXTRACTION





Purpose of Feature Extraction



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

**Extract audio
features**

**Extract text
features.**



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



TOOL

NumPy, SciPy, Matplotlib.



HOW

Computes RMS energy from audio to measure intensity.



WHY

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

AUDIO EXTRACTION





Techniques Used



SPEECH-TO- TEXT



TOOL

SpeechRecognition (VOSK).



HOW

Converts audio to text using offline speech recognition.



WHY

Extracts spoken words for indexing.

Importance of Step 2

DRIVES INDEXING

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

SUPPORTS MULTIMODAL PROCESSING

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

ENABLES SEGMENTATION

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

PREPARES FOR METADATA

Raw features feed into metadata consolidation (Step 4).

