VIDEO INDEXING

STEP2:

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

OBJECT AND FACE DETECTION

TOOL

YOLOv5 (objects), OpenCV (faces).

HOW

Identifies objects (e.g., “dog”) and faces in frames using pre-trained models.

WHY

Tags key visual elements for indexing.

Techniques Used

COLOR AND MOVEMENT DETECTION

TOOL

OpenCV.

HOW

Analyzes frame colors (HSV histograms) and motion (frame differences).

WHY

Captures scene context (e.g., “green park” or “running motion”).

Techniques Used

SPEECH-TO-TEXT

TOOL

SpeechRecognition

HOW

Converts audio to text using Google's speech API.

WHY

Extracts spoken words for text-based search.

Techniques Used

AUDIO EVENT DETECTION

TOOL

Librosa.

HOW

Analyzes audio to classify noise, music, or silence based on energy and spectral features.

WHY

Adds audio context (e.g., “music in a happy scene”).

Techniques Used

ON-SCREEN TEXT DETECTION

TOOL

Tesseract OCR.

HOW

Extracts text from frames (e.g., subtitles).

WHY

Captures displayed text 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).