ITCT IA2:

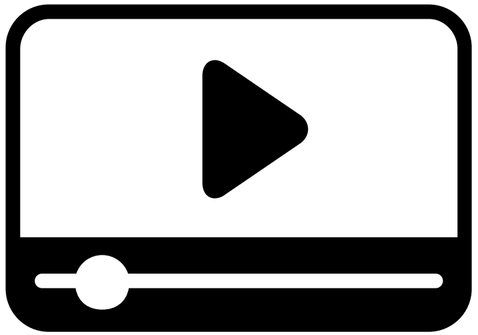
Information Theory in Video Segmentation

By – Ketaki Mahajan, 16014022050

Under the Guidance of – Prof. Makarand Kulkarni

Research Objectives

- Develop novel shot boundary detection methods.
  Identifying where one long continuous shot ends and another begins.
- Create efficient key frame extraction system.
  Selecting important frames, essentially representative frame.
- Improve accuracy of transition detection.
  Identifying visual effects like fades, cuts, dissolves.
- Reduce false detection rate.
  Minimizing errors where the system incorrectly detects shot boundries or transitions.



Video segmentation improves the efficiency, accuracy, and speed of working with videos, especially for summarization, editing, search, and AI-driven video analysis.

Video Transitions

Video Transitions

Abrupt Cut

- Instant transition
- Complete content change

Fades

- Fade-out: gradual darkening to black
- Fade-in: black to full image

(a) (b)

Fig. 1. Consecutive frames from "news" video sequence showing: (a) fade-out and (b) fade-in.

Technical Foundation

- MI measures the mutual dependence between two variables/events.

- Mathematical expression for **Mutual Information** is,

$$MI(X,Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

  where X & Y are two frames, p(x, y) is joint probability distribution of X & Y, P(x) and P(Y) is marginal probability distribution of X & Y

- **Application in video:**
  - Compare consecutive frames to detect significant changes.
  - Identify key frames by measuring information content.
  - Detect scene boundaries when MI between frames drops significantly.
  - Eliminate redundant frames that contain similar information.

Technical Foundation

- **Joint Entropy** measures the total amount of uncertainty or information contained in both frames X and Y.

- Mathematical expression for joint entropy is,

$$H(X,Y) = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log(p(x,y))$$

where H(X,Y) is joint entropy, P(x, y) is joint probability distribution of pixel values of frames X & Y

Joint Entropy captures the combined uncertainty or complexity of both frames together

Mutual Information quantifies how much knowing the pixel values in one frame reduces the uncertainty in the next frame.

Abrupt Cut Detections

- Mutual Information is used for detecting abrupt cuts in video, e.g. where one shot suddenly changes to another.
- MI is used to analyze the similarity between consecutive frames.
- **MI-based detection process:**
  - Calculate MI between consecutive frames.
  - Apply adaptive thresholding.
  - Detect significant drops.

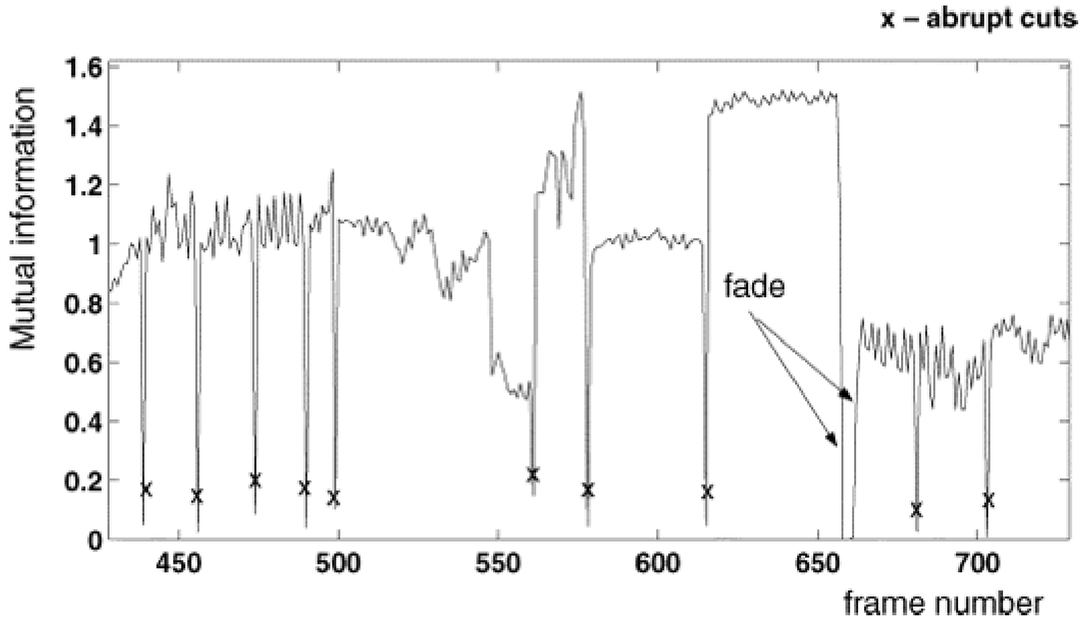


Fig. 2. Time series of the MI from "ABC news" video sequence showing abrupt cuts and one fade.

Fade Detections

- Joint Entropy is used to analyze the total information content in consecutive frames.
- Fade is gradual transition between scenes either by brightening or darkening.
- **JE-based detection process:**
  - Track Joint Entropy values.
  - Identify potential fade regions.
  - Verify fade patterns.

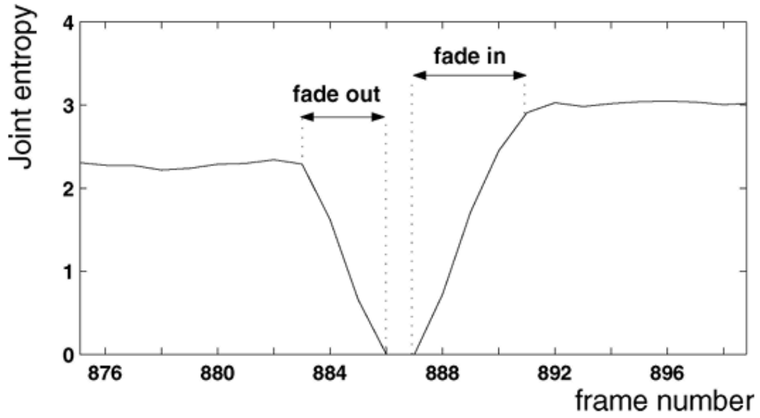


Fig. 3. JE signal from "CNN news" video sequence showing a fade-out and fade-in to the next shot.

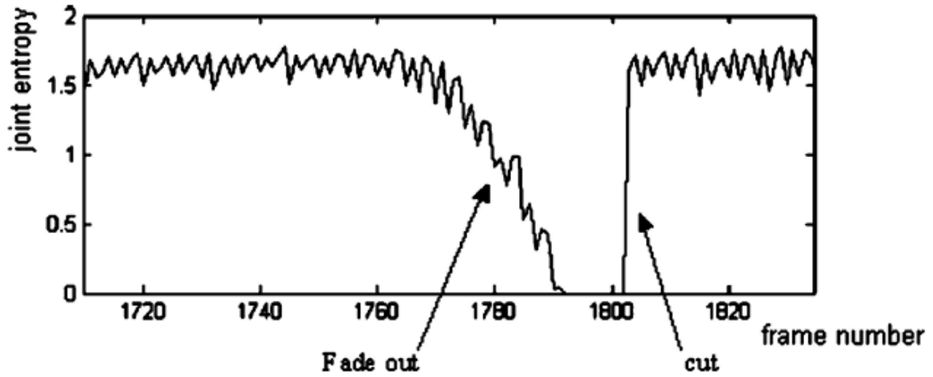
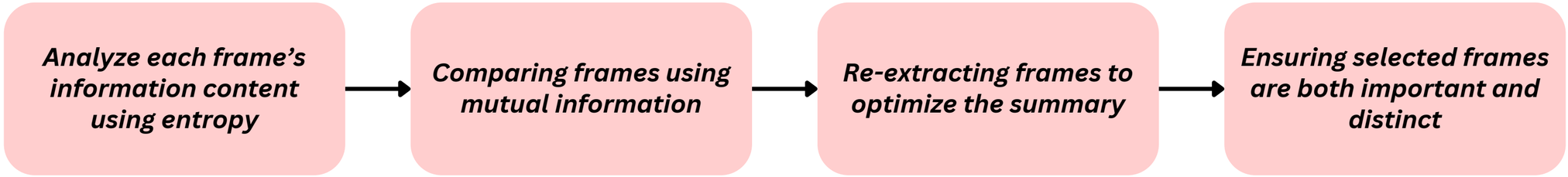


Fig. 4. JE signal from "basketball" video sequence showing a fade-out and a transition from a black frame to the next shot.

Key Frames Selection

- A video is a combination or **group of sequences of images** to form a moving picture.
- The extraction of key frames from the original video uses an innovative **Split-Merge** clustering approach based on mutual information.
- Each shot is analyzed and split into clusters based on significant changes using MI, when MI drops significantly it suggests change in content. The idea is to split video into sections where there are significant visual differences.
- After the initial splits, the system reduces redundancy by merging similar frames, i.e., frames with high MI.



Technical process of key frame extraction

Key Frames Selection

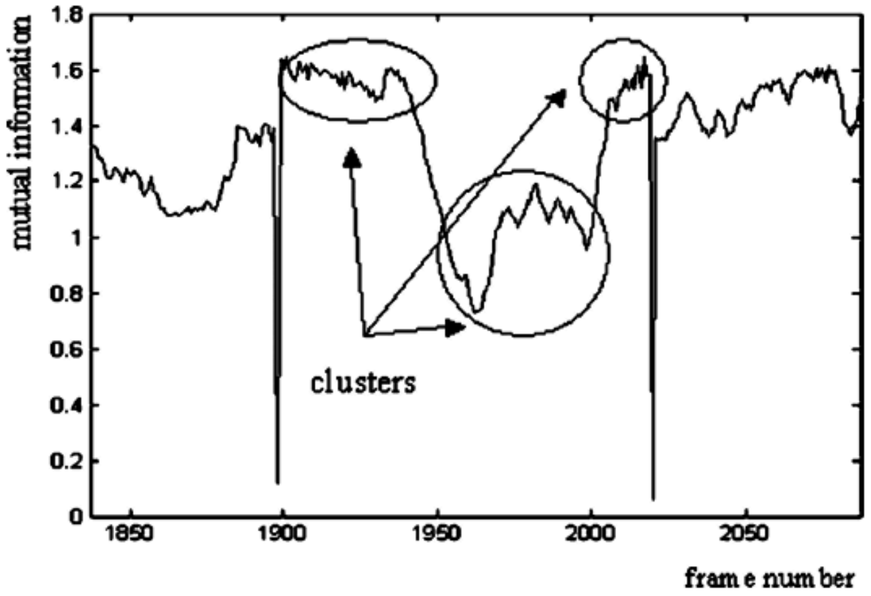


Fig. 5. MI signal from "star" video sequence presenting the clusters created by split-merge method. The selected potential key frames from each cluster is shown in Fig. 6.

Fig. 6. Potential key frames from "star" video sequence extracted from each cluster of the shot.

Experimental Results

TABLE I
VIDEO SET USED IN OUR EXPERIMENTS

video	frames	cuts	fade-ins	fade-outs
basketball	3882	44	7	4
news	9446	40	6	6
football	5589	28	0	0
movie	19722	147	0	0
TREC video sequences				
6 debate videos	125977	230	0	0
4 CNN news videos	209978	1287	57	57
4 ABC news videos	206144	1269	64	69

Proposed method was tested on several TV sequences containing many commercials

TABLE II
FIXED THRESHOLD SHOT CUT DETECTION RESULTS

	MI method		**Combined histogram method**		**Color histogram method**	
video	Recall	Precision	Recall	Precision	Recall	Precision
basketball	**1.00**	**1.00**	0.91	0.97	0.52	0.85
news	0.96	**1.00**	0.96	0.98	0.80	0.89
football	0.93	1.00	0.96	1.00	0.68	0.68
movie	**1.00**	**1.00**	0.93	0.98	0.87	0.93
6 debate videos	1.00	0.99	1.00	1.00	0.87	1.00
4 CNN news	**0.96**	**0.96**	0.87	0.83	0.85	0.84
4 ABC news	**0.97**	**0.94**	0.85	0.81	0.87	0.73
TREC total	**0.97**	**0.95**	0.87	0.83	0.86	0.80

Proposed method results comparison with other traditional methods. The best reported abrupt cut detection results for recall and precision are 93% and 92%, respectively, whereas method proposed produces 97% recall and 95% precision.

Experimental Results

Analysis of football game, specifically focusing on detecting and analyzing camera flashes and video transitions. Graph(a) shows histogram measurements and Graph(b) shows MI calculated values.

Mathod(a) shows more noise/fluctuations in signal, pattern is less consistent adn harder to distinguish between different types of events.

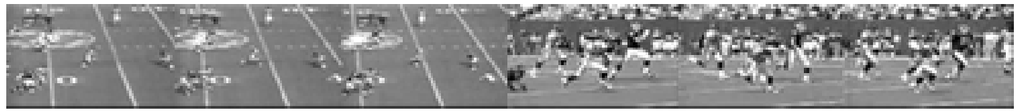


Fig. 8. Consecutive frames from "football" video sequence showing an abrupt cut between two shots coupled with large video object motion.

Fig. 9. Consecutive frames from "football" video sequence showing a big object appearance in front of the camera during panning.

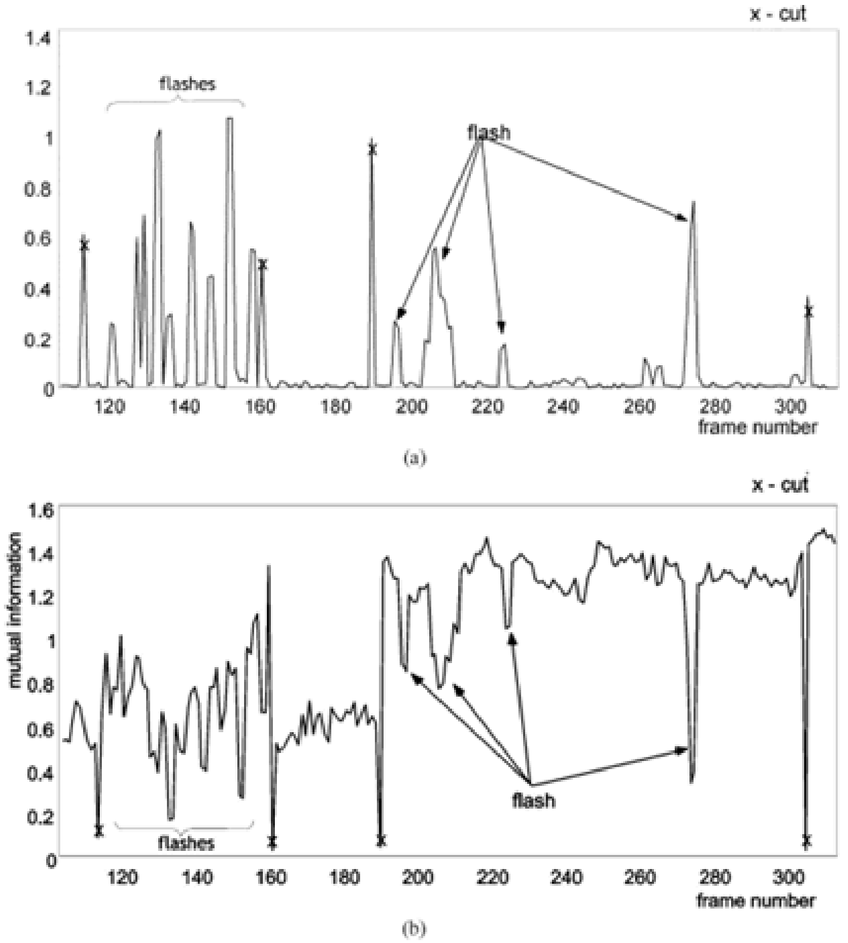


Fig. 10. Part of video sequence containing many camera flashes. (a) Color histogram comparison and (b) MI calculated for the same part of video sequence.

Advantages

The system offers several key advantages over traditional methods:

- Robust to camera flashes
- Handles motion well
- Lower false detection rates
- Works with short shots
- Accurate fade detection

Limitations

Challenges with similar scenes & issues with partial transitions

If two frames have similar lighting and colors, the algorithm may not register a change, leading to missed detections. This scenario is illustrated in Figure 12(a).

Sometimes, a shot change may occur only in a specific area of the video frame, rather than affecting the entire frame. If the algorithm focuses on overall frame characteristics, it might overlook these localized changes. This issue is represented in Figure 12(b).

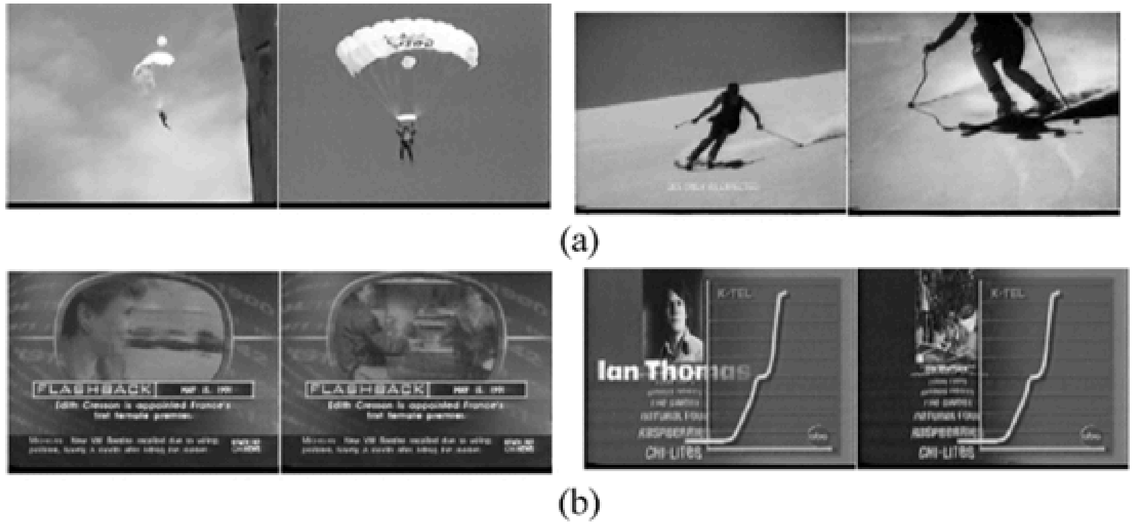


Fig. 12. Consecutive frames from video sequences presenting abrupt cuts, which caused missed shot cut detection. (a) Shot changes between two images with very similar spatial color distribution and (b) shot change occurs only in a part of the video frame.

Limitations

Artistic camera edit detection & noisy black frame detection

Artistic camera edits, such as those often used in commercials, can also result in false detections.

These edits may intentionally manipulate color and composition, confusing the detection algorithm.

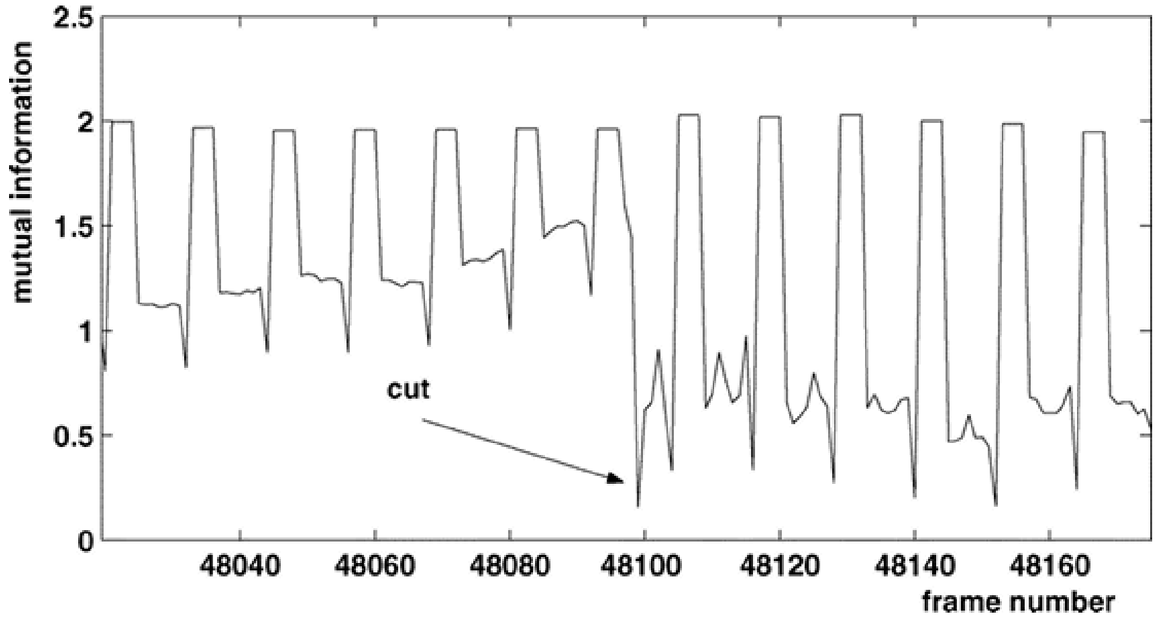


Fig. 11. MI of temporally subsampled video sequence.

References

Primary Research Paper:
https://ieeexplore.ieee.org/abstract/document/1564125

https://link.springer.com/chapter/10.1007/978-1-84800-007-0_14

https://quantdare.com/what-is-mutual-information/

https://www.inf.ed.ac.uk/teaching/courses/fmcs1/slides/lecture25.pdf

Thank You!