Video Syntax Analysis

Wei-Ta Chu

Shot Boundary Detection / Shot Change Detection

- Shot
  - A basic unit for advanced accessing – browsing, summarization, retrieval
- Keyframes
  - Representative frame(s) of a shot
- Issues
  - Large camera/object motion
  - Editing effects: dissolve, wipe, fade
  - Flashlight

Types of Shot Change

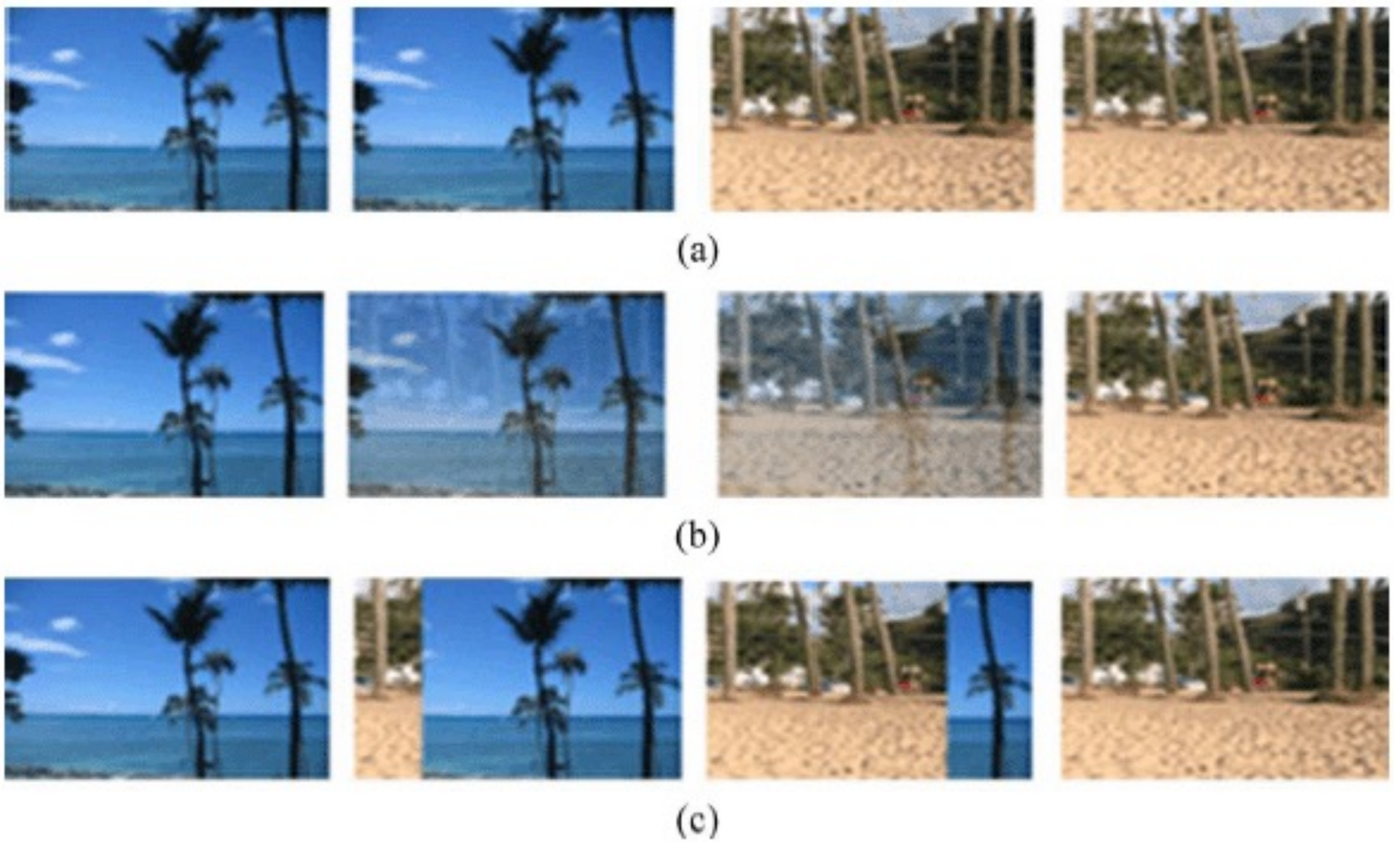
- Abrupt change (hard cut)
  - Cut occurs in a single frame when stopping and restarting the camera
- Gradual transition
  - Fade-in: gradual increase in intensity starting from a black frame
  - Fade-out: gradual decrease in intensity resulting a black frame
  - Dissolve: transiting from the end of one clip to the beginning of another
  - Wipe: One image is replaced by another with a distinct edge that forms a shape.
  - …

Examples of Shot Changes

Cut

Dissolve

Wipe



Li and Lee. “Effective detection of various wipe transitions” IEEE Trans. on Circuits and Systems for Video Technology, vol. 17, no. 6, pp. 663-673, 2007.

Examples of Fade

Fade out

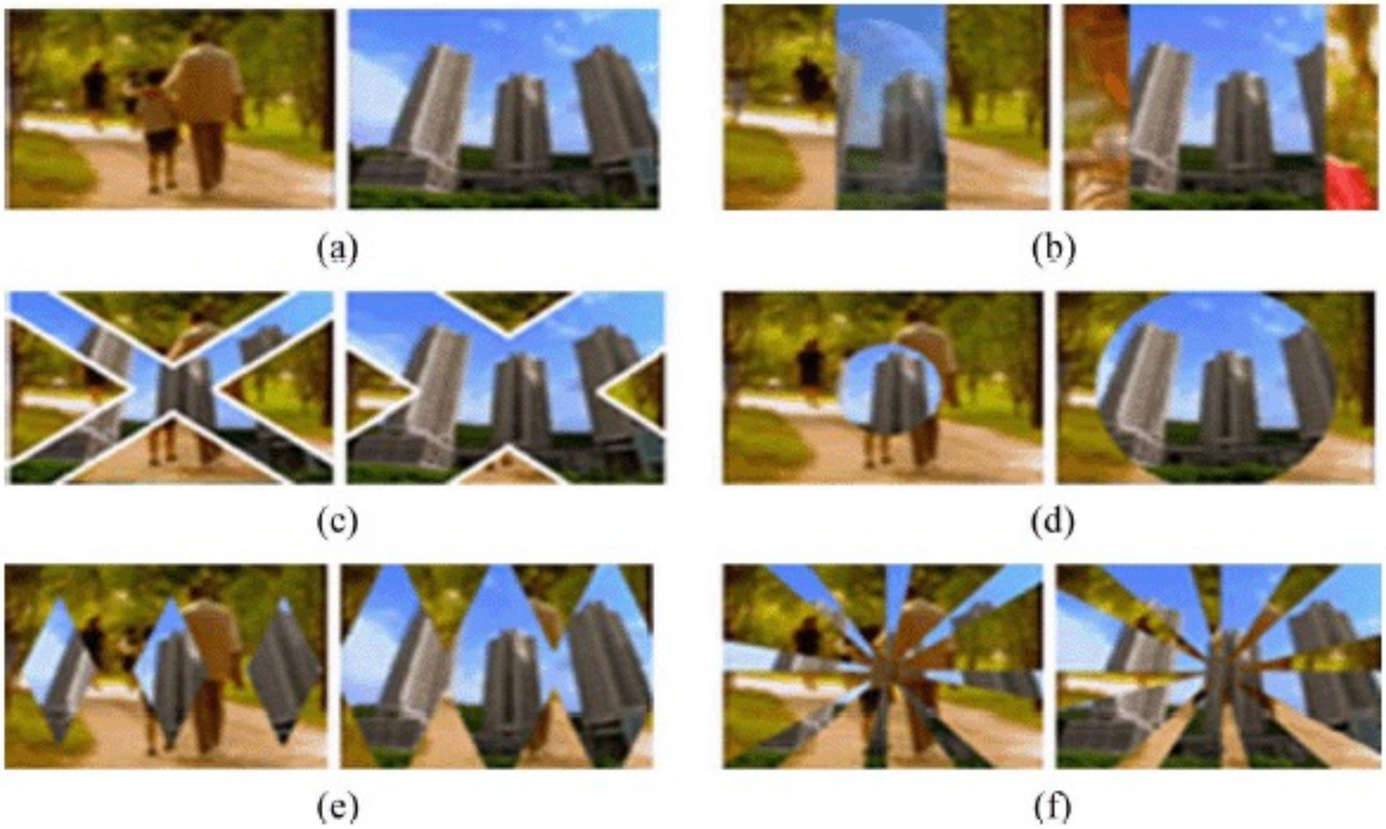
(a)

Fade in

(b)

Cernekova, et al., “Information theory-based shot cut/fade detection and video summarization” IEEE Trans. on Circuits and Systems for Video Technology, vol. 16, no. 1, pp. 82-91, 2006.

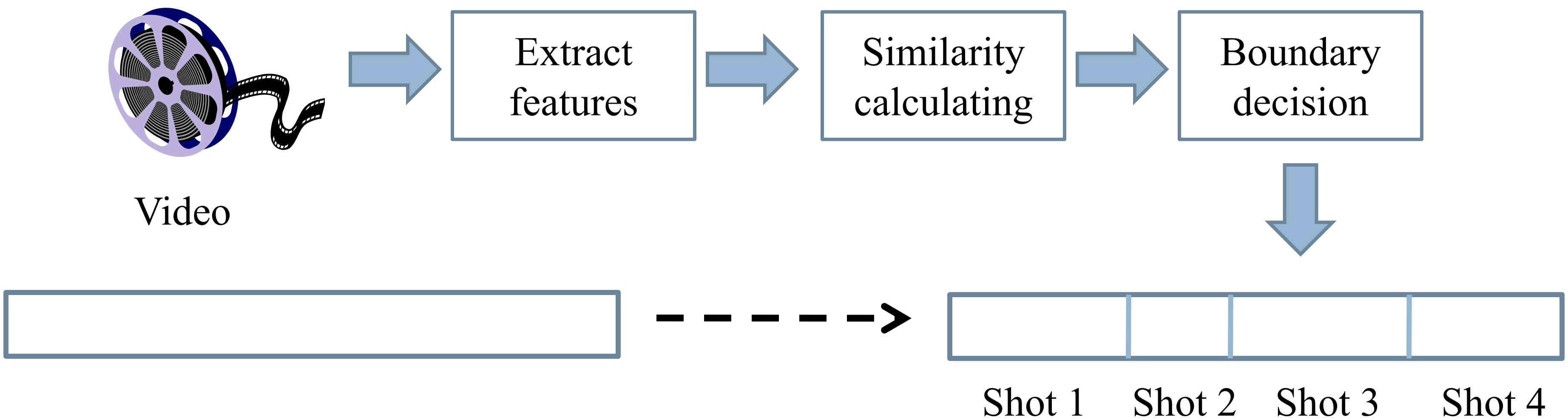
Different Types of Wipe



Li and Lee. “Effective detection of various wipe transitions” IEEE Trans. on Circuits and Systems for Video Technology, vol. 17, no. 6, pp. 663-673, 2007.

Video example: http://en.wikipedia.org/wiki/Wipe_%28transition%29

Detection Process



Histogram Comparison

- Less sensitive to object motion, since it ignores the spatial changes in a frame.

$$SD_i = \sum_{j=1}^{G} |H_i(j) - H_{i+1}(j)| \; .$$

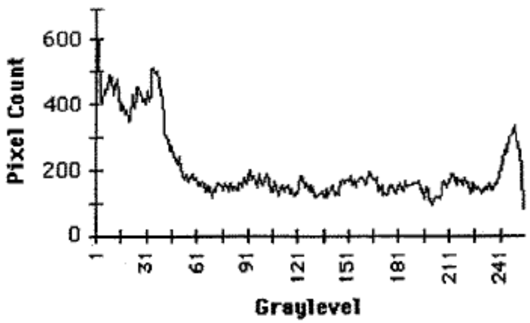
$H_i(j)$: the histogram value for the ith frame, where j is one of the G grey levels.

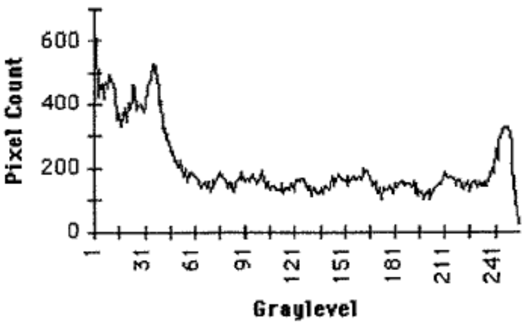
$$SD_i = \sum_{j=1}^{G} \frac{|H_i(j) - H_{i+1}(j)|^2}{H_{i+1}(j)}$$

χ^2-test equation makes the histogram comparison reflect the difference between two frames more strongly.

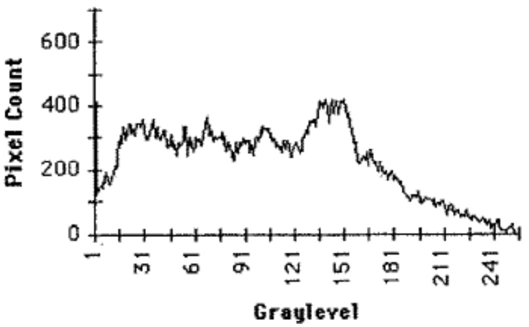
Histogram Comparison – Example

Example video sequence





The intensity histogram
of the first three frames



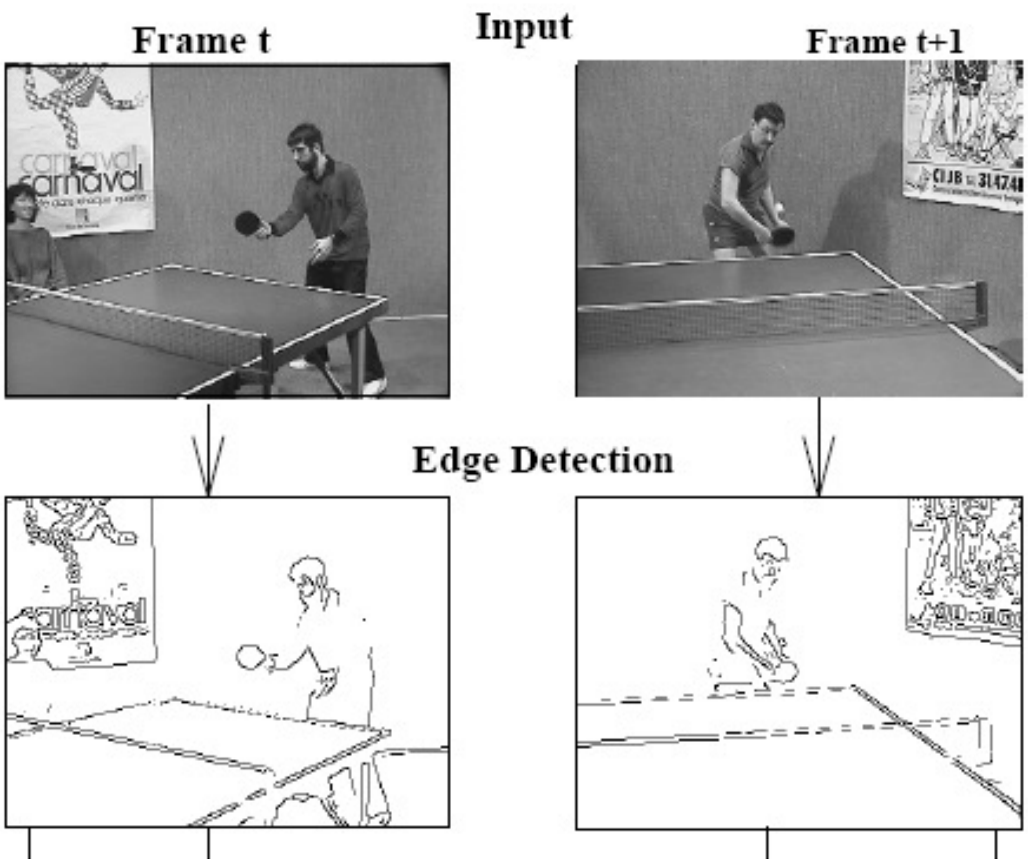
Histogram Comparison

- Color histogram difference

$$CHD_i = \frac{1}{N} \cdot \sum_{r=0}^{2^B-1} \sum_{g=0}^{2^B-1} \sum_{b=0}^{2^B-1} \left| p_i(r,g,b) - p_{i-1}(r,g,b) \right|$$

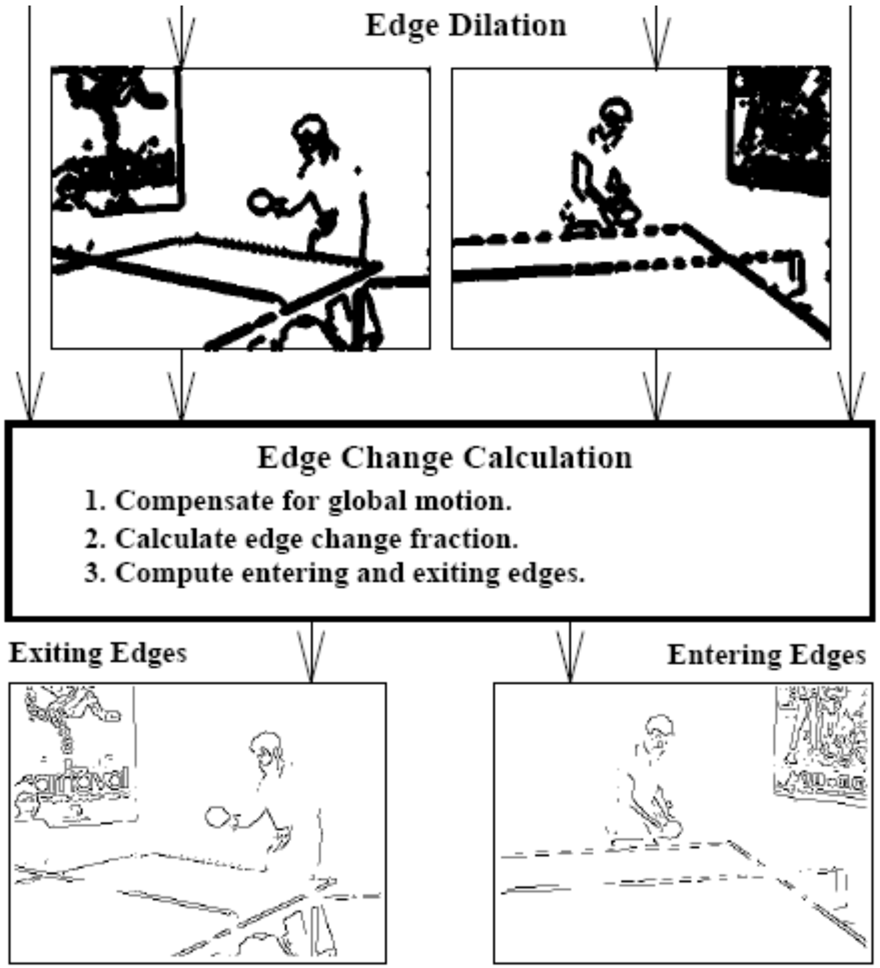
$p_i(r,g,b)$ is the number of pixels of color (r,g,b) in frame I_i of N pixels.
Each color component is discritized to 2^B different values.

Edge Change Ratio



Zabih, et al., “A feature-based algorithm for detecting and classifying scene breaks” Proc. Of ACM Multimedia, pp. 189-200,1995.

Edge Change Ratio



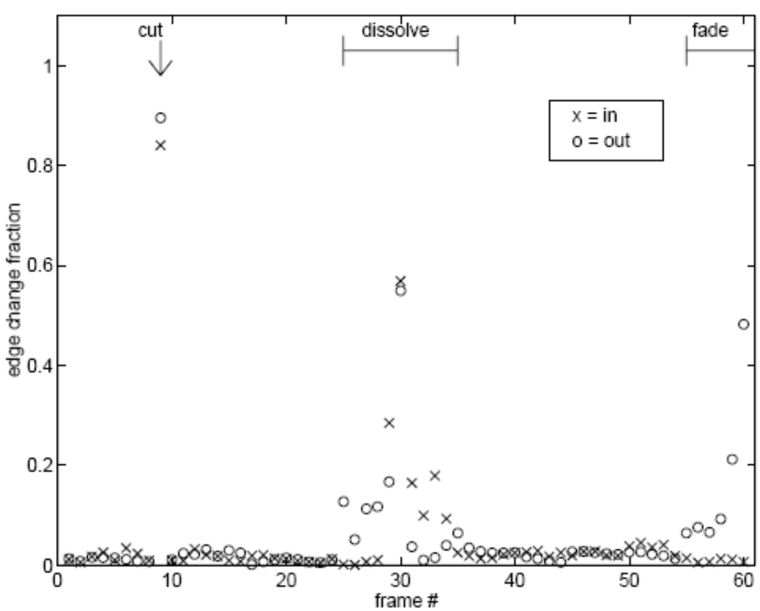
Edge Change Ratio

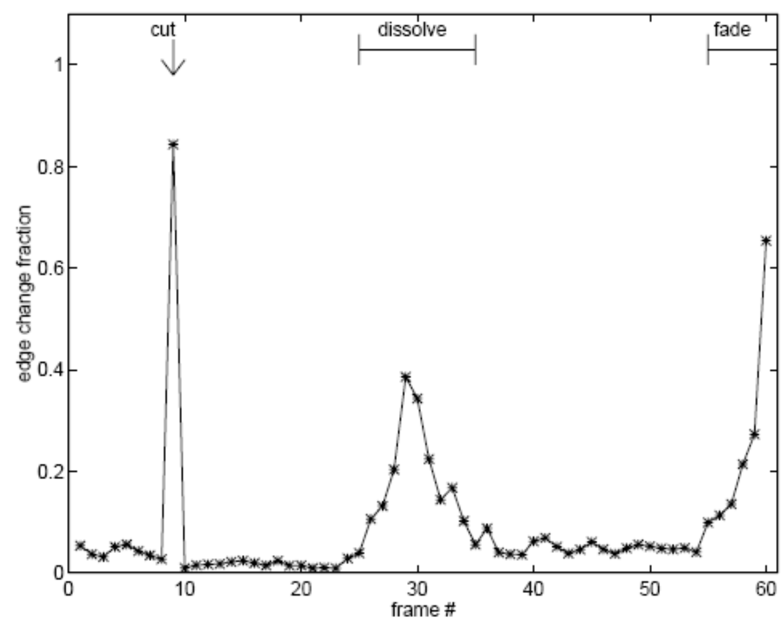
- Edge change ratio

$$ECR_n = max(X_n^{in}/\sigma_n, X_{n-1}^{out}/\sigma_{n-1})$$

σ_n is the number of edge pixels in frame n

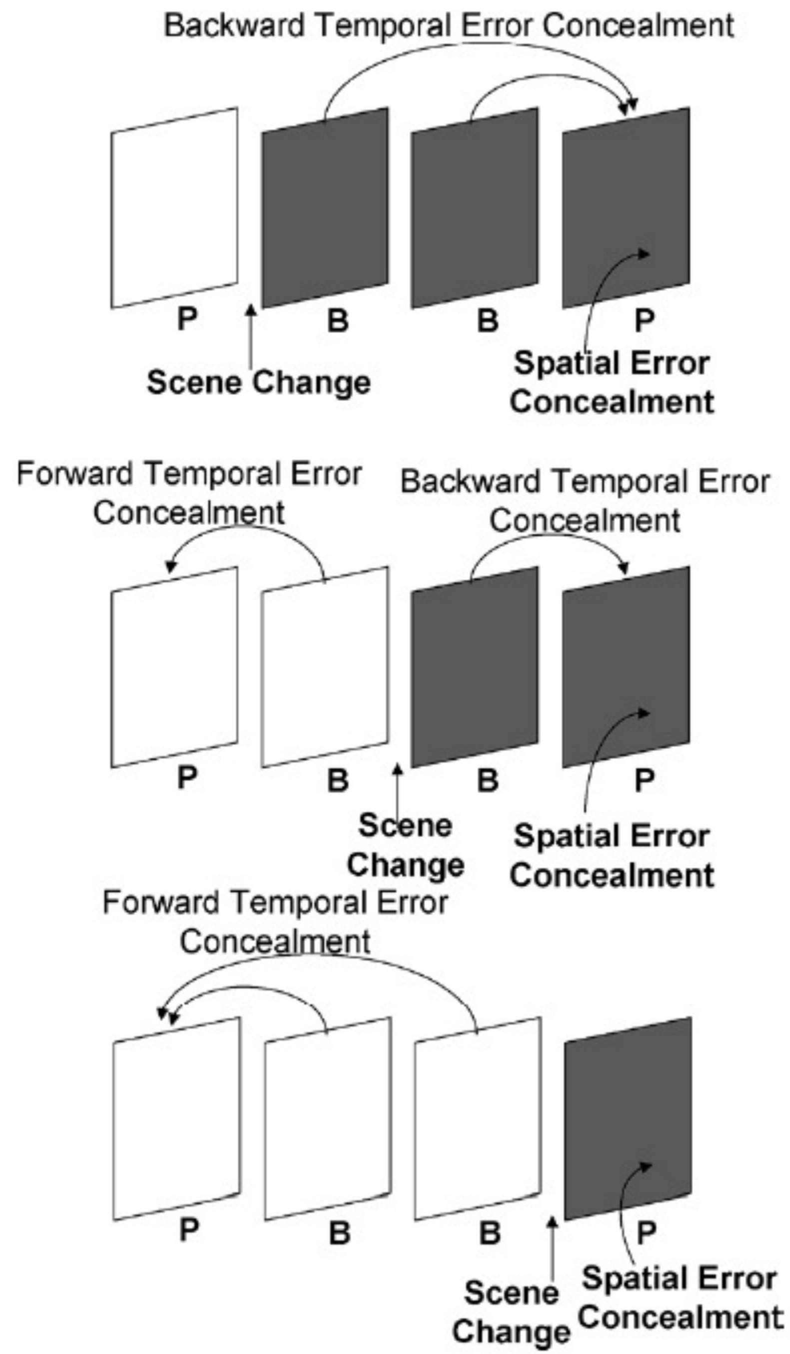
X_n^{in} and X_{n-1}^{out} are the numbers of entering and exiting edge pixels in frames n and $n-1$.





Motion Vectors

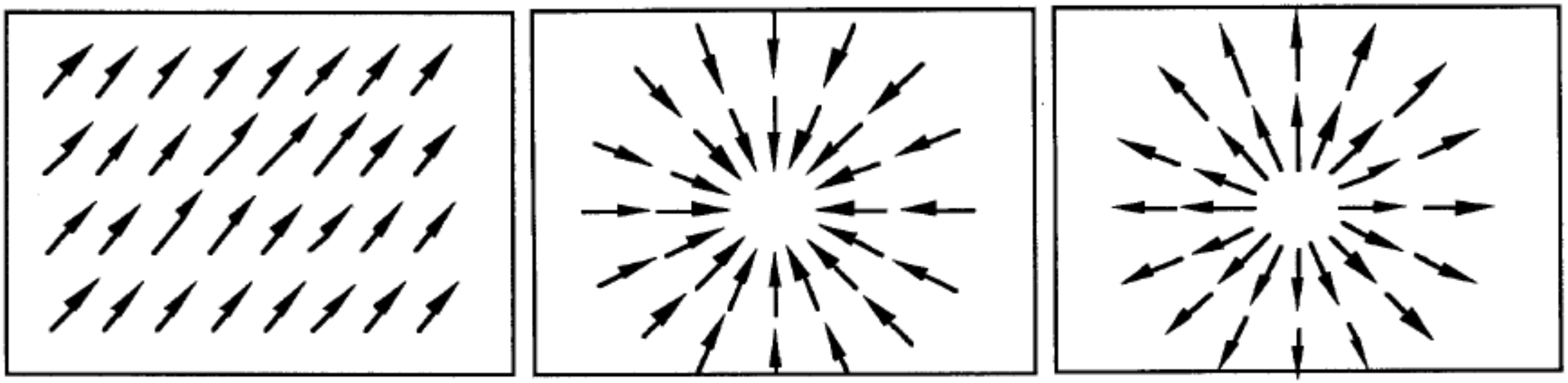
- Using the direction of motion prediction to be the cues for shot change detection



Pei, et al., "Scene-effect detection and insertion MPEG encoding scheme for video browsing and error concealment" IEEE Trans. on Multimedia, vol. 7, no. 4, pp. 606-614, 2005.

Motion Vectors

- Using motion vector information to filter out false positives

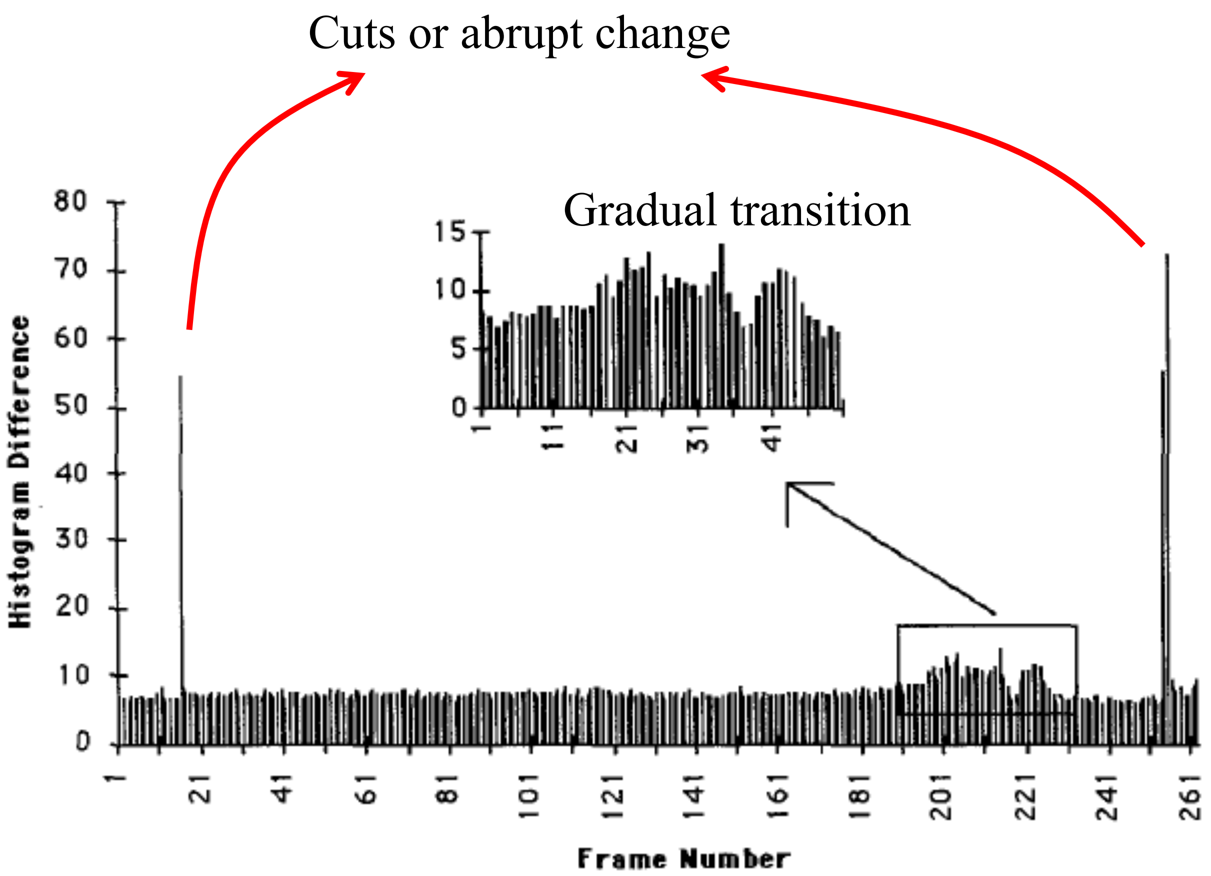


a b c

Fig. 6a–c. Motion vector patterns resulting from camera panning and zooming. **a** ↙ Camera panning direction. **b** Camera zoom-out. **c** Camera zoom-in

Zhang, et al., “Automatic partitioning of full-motion video” Multimedia Systems Journal, vol. 1, pp. 10-28, 1993.

Gradual Transition Detection



1. Twin-Comparison Approach

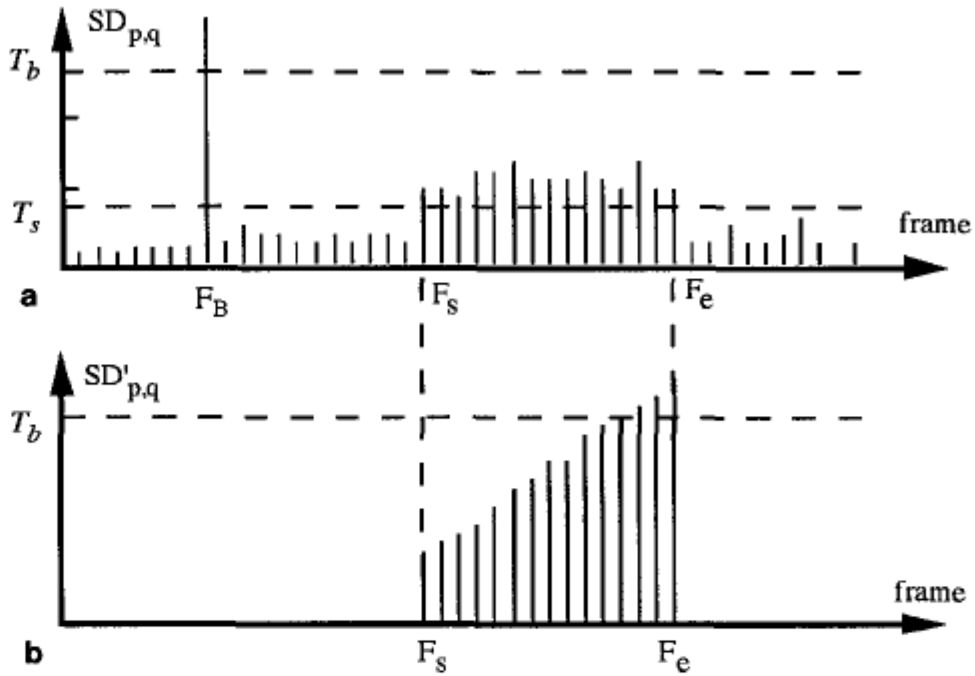
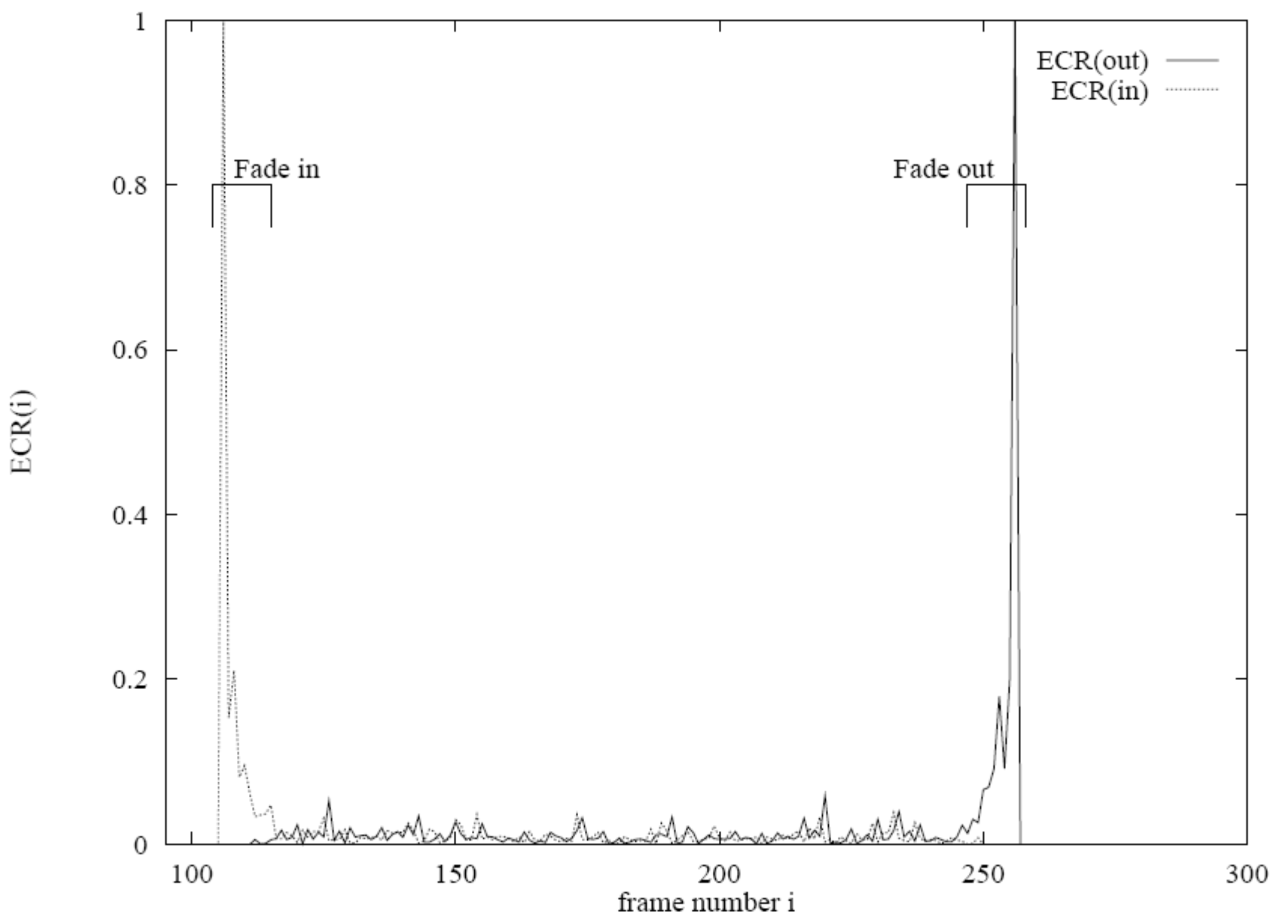


Fig. 5a,b. Illustration of twin-comparison. $SD_{p,q}$, the difference between consecutive frames defined by the difference metric; $SD'_{p,q}$, the accumulated difference between the current frame and the potential starting frame of a transition; T_s, the treshold used to detect the starting frame (F_s) of a transition; T_b, the thresold used to detect the ending frame (F_e) of a transition. T_b is also used to detect camera breaks and F_B is such a camera break. $SD'_{p,q}$ is only calculated when $SD_{p,q} > T_s$

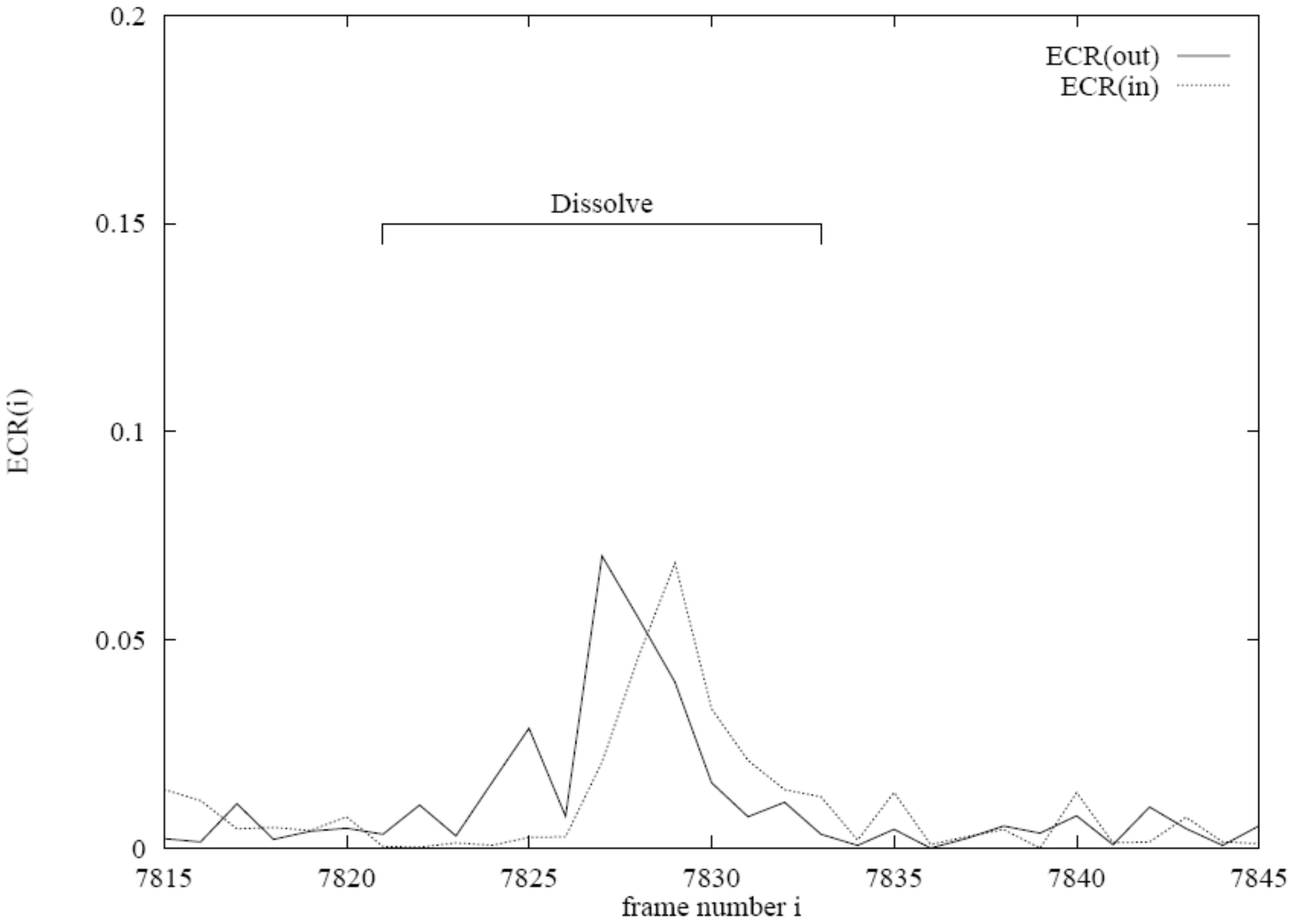
Zhang, et al., "Automatic partitioning of full-motion video" Multimedia Systems Journal, vol. 1, pp. 10-28, 1993.

2. Edge Change Ratio



Lienhart, R., “Comparison of automatic shot boundary detection algorithms” Proc. of SPIE Storage and Retrieval for Image and Video Databases VII, vol. 3656, pp. 290-301, 1999.

2. Edge Change Ratio



3. Characterizing a Wipe Transition

Suppose a wipe transition spreads over N frames. Denote the scene change region between frames t and $t+1$ as ξ_t; S_1 and S_2 represent the current shot and the next shot, respectively. An ideal wipe sequence (i.e., during an ideal wipe transition, the two shots are assumed to be motionless) can be modeled as

$$S(x,y,t) = \begin{cases} S_2(x,y), & \forall (x,y) \in \xi_1 \cup \xi_2 \cup \ldots \cup \xi_{t-1} \\ S_1(x,y), & \text{otherwise} \end{cases}, \quad 1 \le t \le N \tag{1}$$

where $S(x,y,t)$ is the pixel intensity at position (x,y) in frame t. N is the total number of frames in the sequence. ξ_t is computed as

$$\xi_t = \{(x,y) | S(x,y,t) \neq S(x,y,t+1\} . \tag{2}$$

Evaluation

- Precision
  - The percentage of retrieved items that are desired items
- Recall
  - The percentage of desired items that are retrieved.

$$\text{Precision} = \frac{\#\ \text{Correctly retrieved items}}{\#\ \text{All retrieved items}} = \frac{\#\ \text{Correctly retrieved items}}{\#\ \text{Correctly retrieved items} + \#\ \text{Falsely retrieved items}}$$

$$\text{Recall} = \frac{\#\ \text{Correctly retrieved items}}{\#\ \text{All relevant items}} = \frac{\#\ \text{Correctly retrieved items}}{\#\ \text{Correctly retrieved items} + \#\ \text{Items that are not retrieved}}$$

Evaluation – Other Terms

- Miss
  - # Items that are not retrieved
- True positive (TP)
  - # Correctly retrieved items
- False positive (FP)
  - # Falsely retrieved items
- True negative (TN)
  - # Correctly missed items
- False negative (FN)
  - # Items that are not retrieved

	Actual positive	Actual negative
Predicted positive	TP	FP
Predicted negative	FN	TN

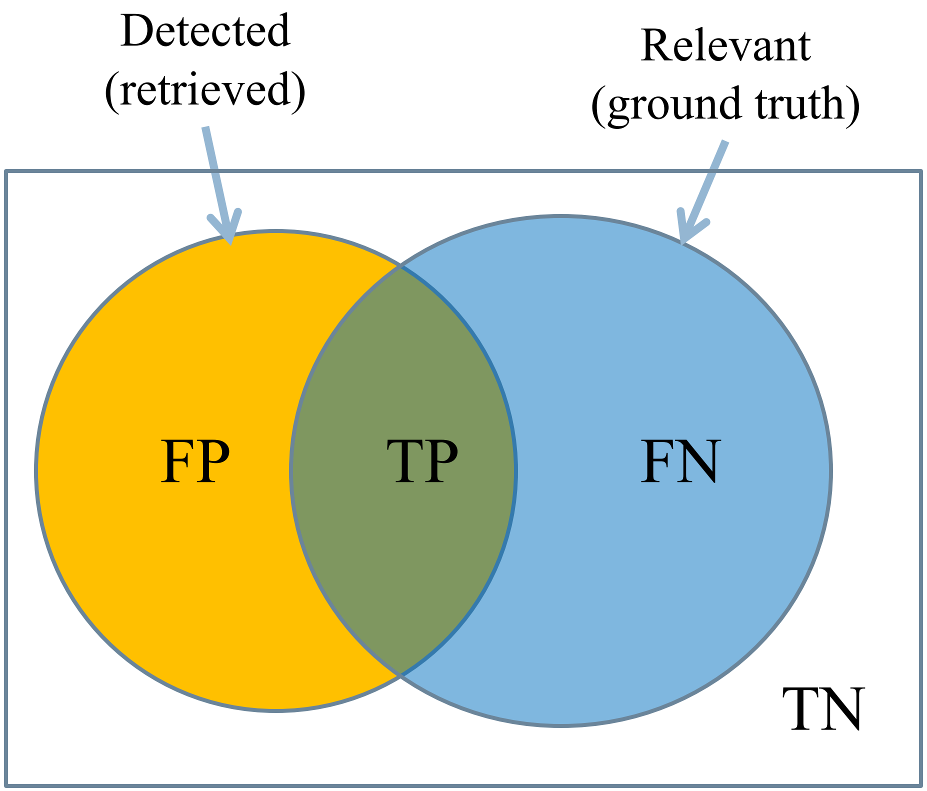
$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{Truth positive rate} = \frac{TP}{TP+FN}$$

$$\text{False positive rate} = \frac{FP}{FP+TN}$$

Evaluation



	Actual positive	Actual negative
Predicted positive	TP	FP
Predicted negative	FN	TN

$$\text{Precision} = \frac{TP}{TP+FP}$$

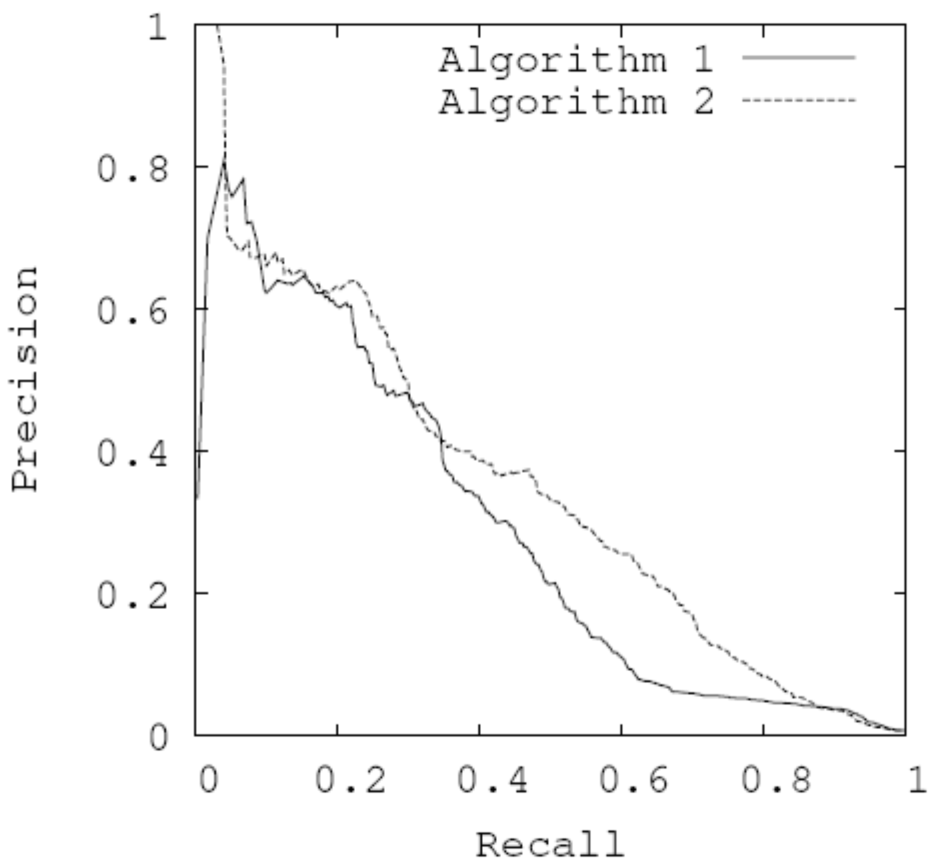
$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{Truth positive rate} = \frac{TP}{TP+FN}$$

$$\text{False positive rate} = \frac{FP}{FP+TN}$$

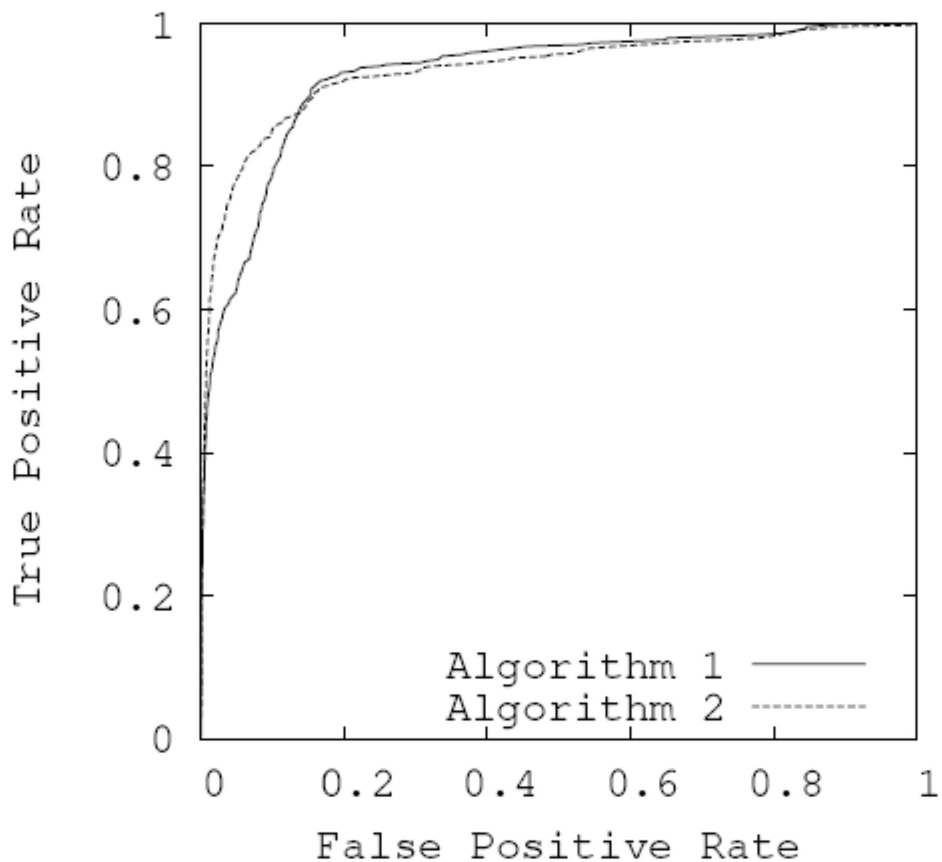
Relationship between Precision & Recall

- Precision-Recall (PR) curve



Relationship between True Positive and False Positive

- Receiver Operator Characteristic (ROC) curve



Using PR or ROC Curves?

- ROC curves can present an overly optimistic view of an algorithm's performance if there is a large skew in the class distribution.
- Number of true negative examples greatly exceeds the number of positive examples. Thus a large change in the number in false positives can lead to a small change in the false positive rate.
- Precision compares false positives to true positives and better captures the algorithm's performance.

Davis, et al., "The relationship between precision-recall and ROC curves" Proc. of International Conference on Machine Learning, pp. 233-240, 2006.

Comparison of Shot Boundary Detection Techniques

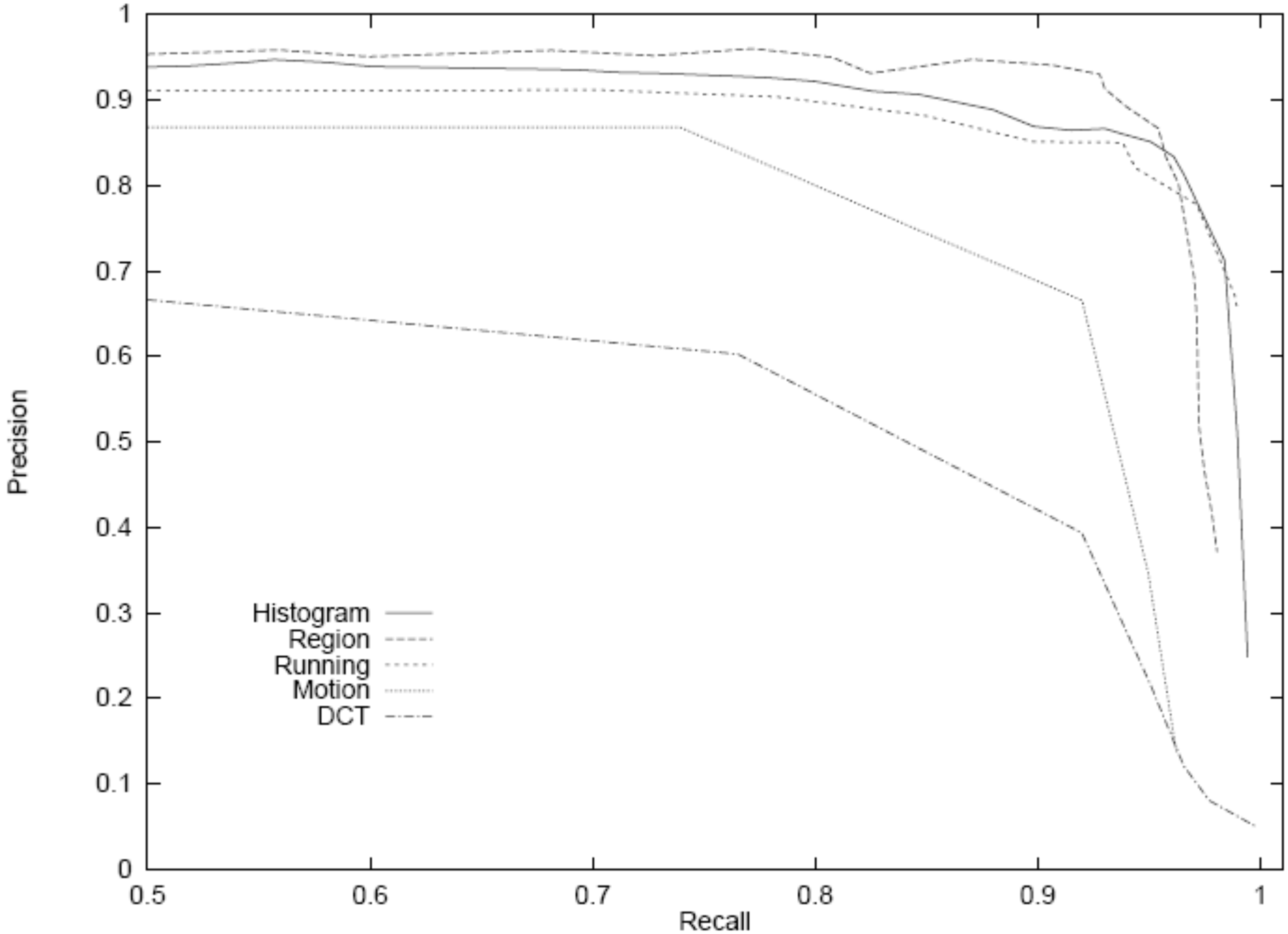
- Methods
  - Histograms, region histograms, running histograms, motion-compensated pixel differences, DCT coefficient differences
- Evaluation data

Video type	# Frames	Cuts	Gradual transitions
TV	133204	831	42
News	81595	293	99
Movie	142507	564	95
Commercial	51733	755	254
Misc.	10706	64	16
Total	419745	2507	506

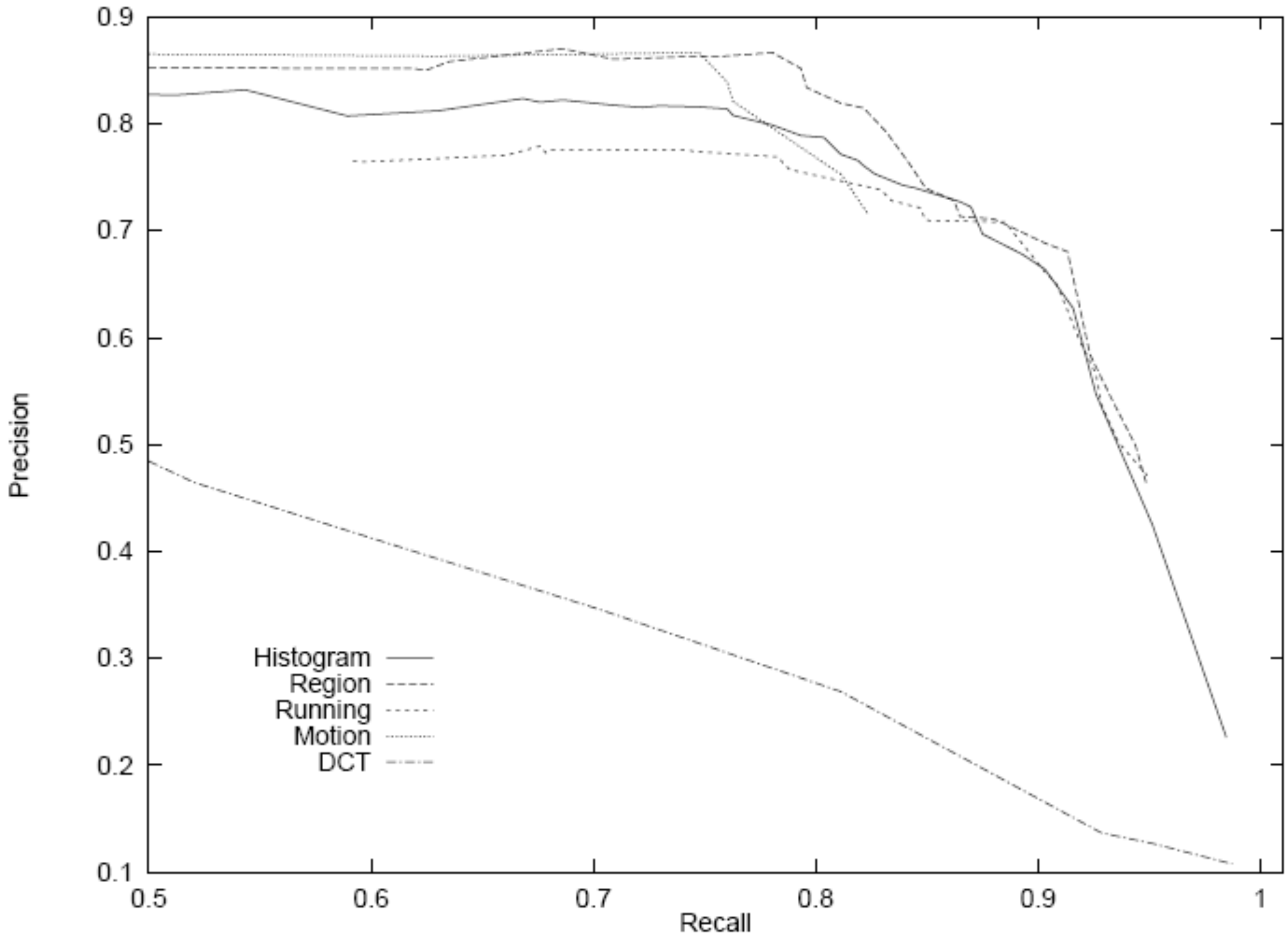
Methods Compared

- Histogram (64-bin gray-level) difference, single threshold
- Region (block) histogram
  - 16 blocks, 64 gray-scale histograms, difference threshold for each block, and count threshold for changed blocks
- Running histogram (Twin method)
  - 64 gray-scale histogram for each frame, twin thresholds
  - Compute motion vectors. If excessive motion, reject gradual changes
- Motion compensated pixel difference
  - 12 blocks per frame, motion vector for each block
  - Compute average residual errors, if larger than high threshold, detected as a cut
  - Use cumulative errors to detect gradual changes (similar to above)
  - Use motion vectors to reject false gradual changes
- DCT difference
  - Concatenate 15 coefficients of same locations from different blocks to form a vector
  - Compute (1-inner product of two vectors from consecutive frames)

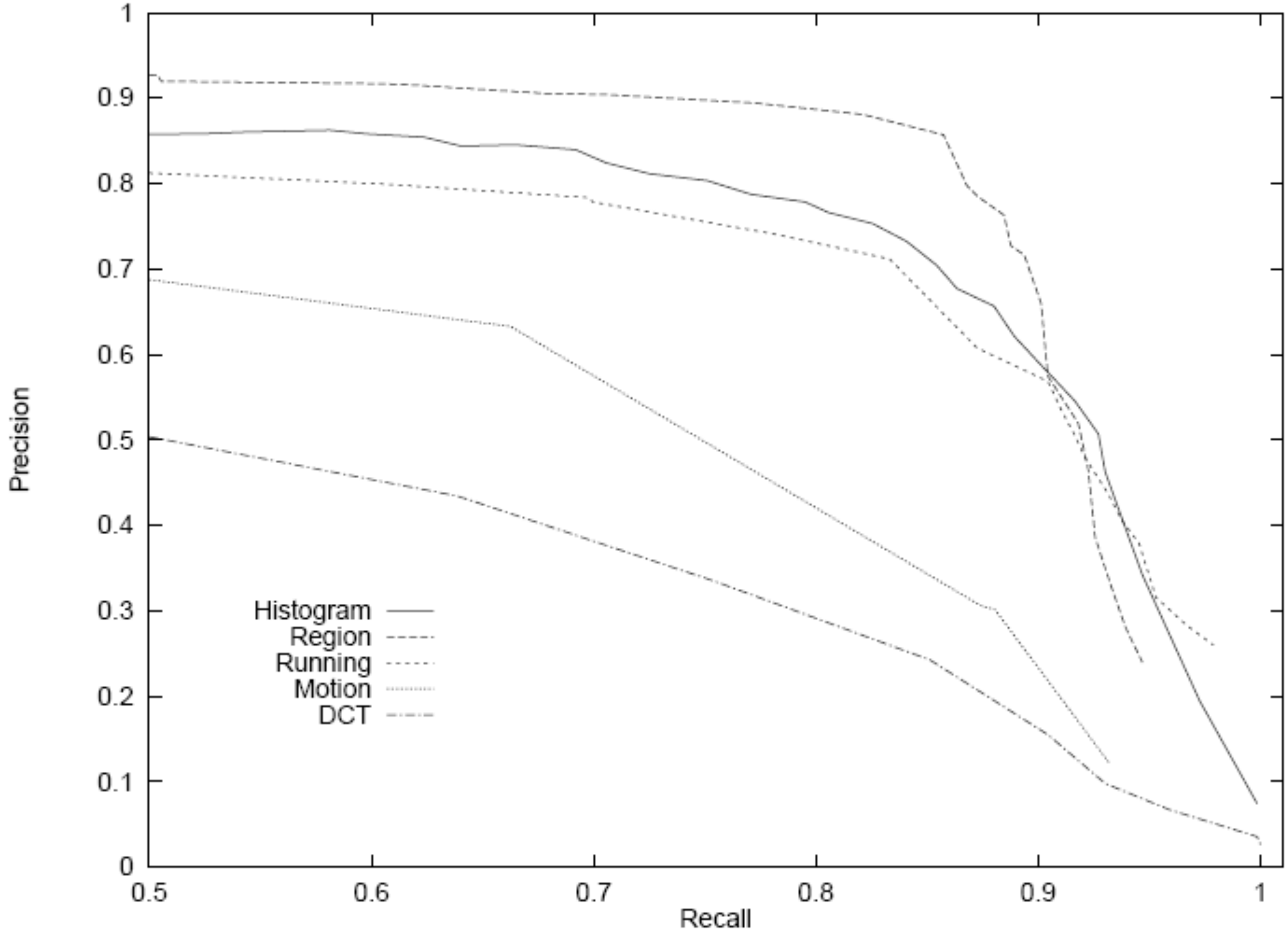
PR Curve for TV program



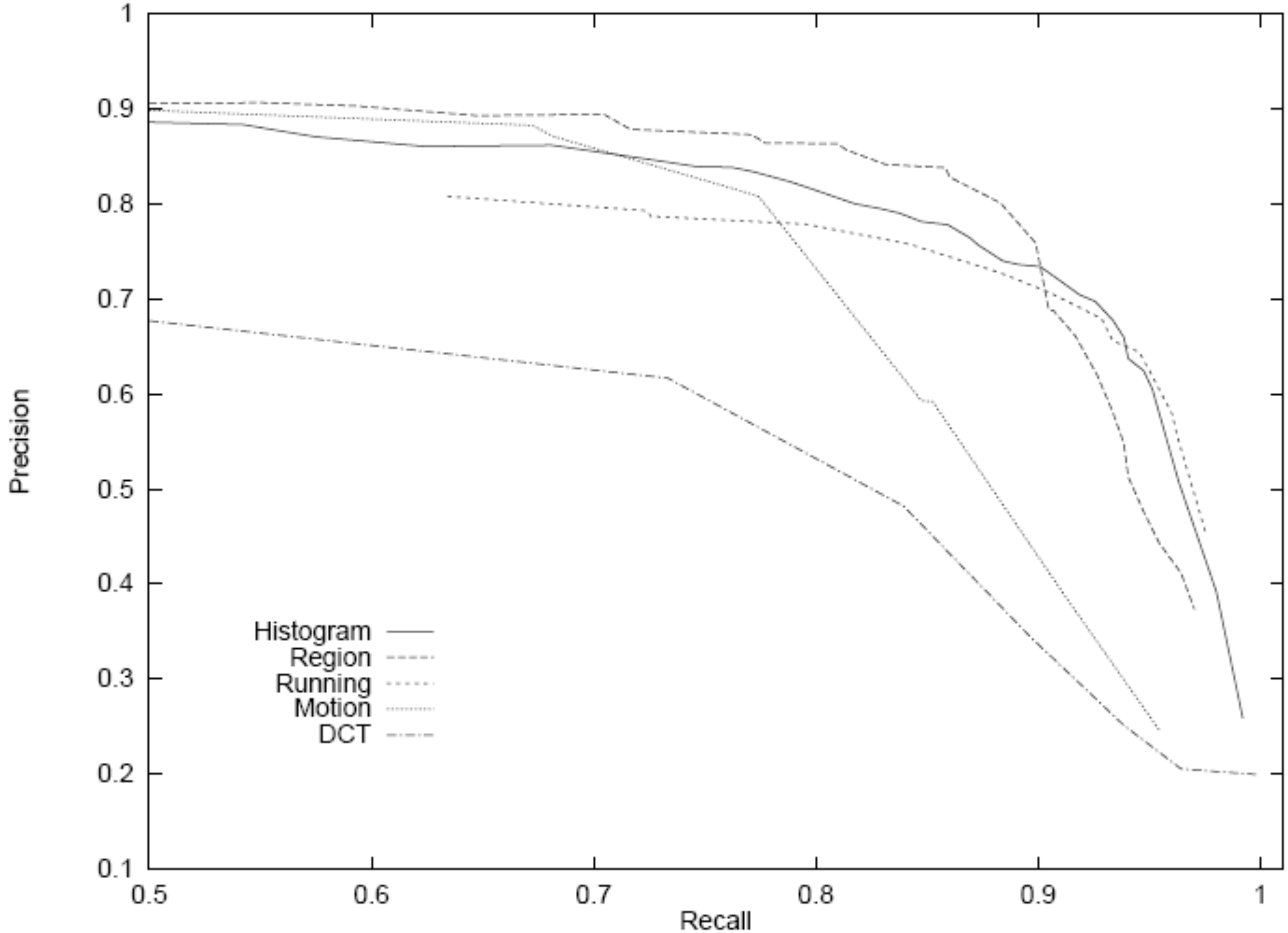
PR Curve for News program



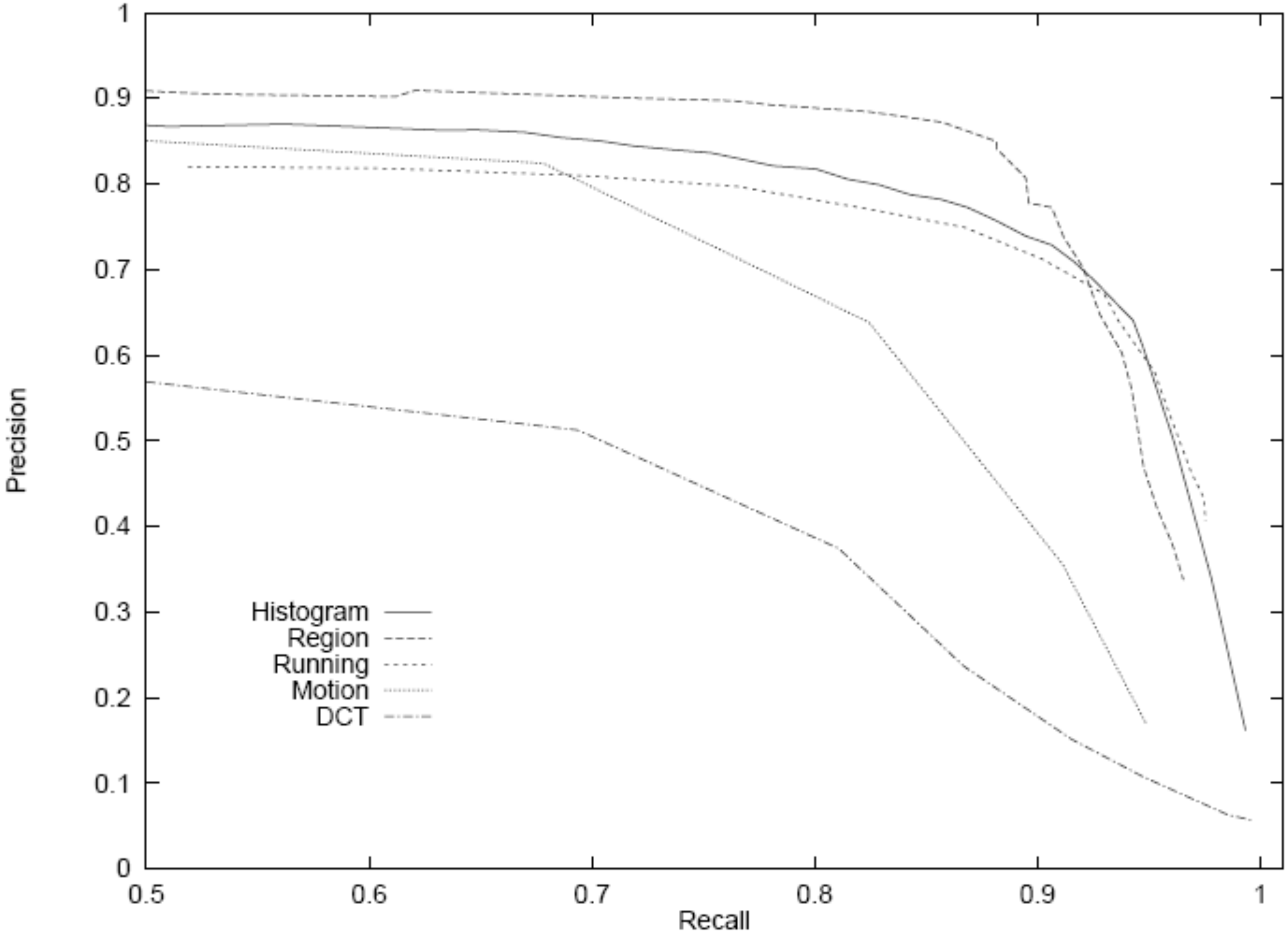
PR Curve for Movie Videos



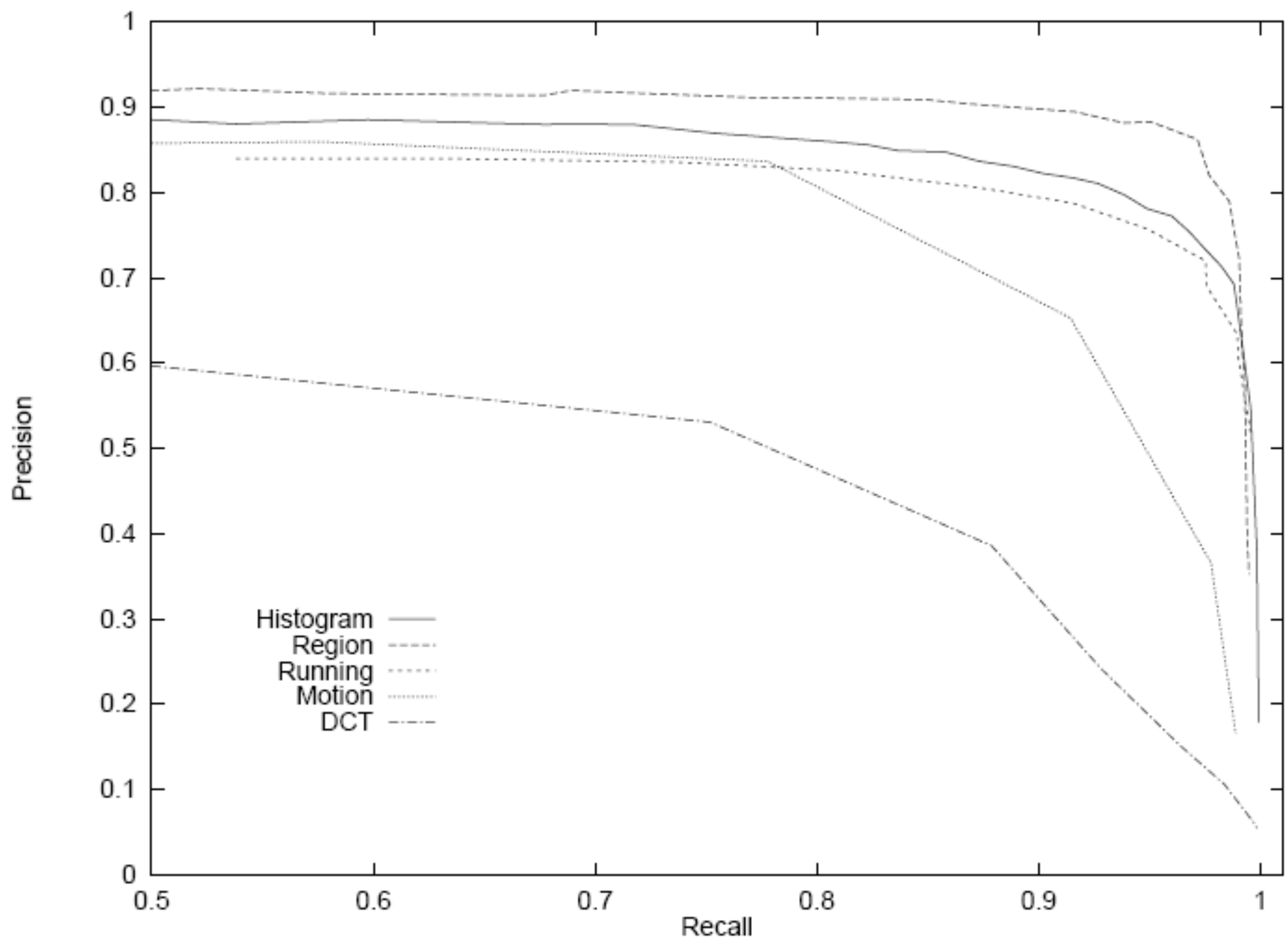
PR Curve for Commercials



PR Curve for All Data



PR Curve for All Data – Cut Only



Observations

- Histogram-based method is consistent
  - Produced the first or second best precision
  - Simplicity & straightforward
- Region algorithm seems to be the best
  - Where recall is not the highest priority
- Running algorithm seems to be the best
  - Where recall is important
  - Motion vector is helpful to reduce false positives
- DCT the worst
  - Large number of false positives in black frames

References

- J.S. Boreczky, et al., "Comparison of video shot boundary detection techniques" Proc. of SPIE Conference on Storage and Retrieval for Image and Video Databases, vol. 2670, 1996. **(must read)**
- R. Lienhart, "Comparison of automatic shot boundary detection algorithms" Proc. of SPIE Storage and Retrieval for Image and Video Databases VII, vol. 3656, pp. 290-301, 1999.
- J. Yuan, et al., "A formal study of shot boundary detection" IEEE Trans. on Circuits and Systems for Video Technology, vol. 17, no. 2, pp. 168-186, 2007.
- A. Hanjalic, "Shot-boundary detection: unraveled or resolved?" IEEE Trans. on Circuits and Systems for Video Technology, vol. 12, no. 2, pp. 90-105, 2002.