# Reliable Dissolve Detection

Multimedia Signal Processing and Protection Group

Academia Sinica, Taiwan

# Outline

- Motivation.
- Introduction.
- Related Work.
- A New Reliable Method.
- Result & Future Work.

# Motivation

- Shot segmentation is one of the major issues in the video indexing/abstraction problem.
- Detecting abrupt change is easy, but not a single algorithm is feasible for all gradual transitions.
- Dissolve effect, among the various gradual transitions, is hardest to be detected due to potential mixture with local and global motions.

# Introduction (1/5)

- **What is a shot ?**

  - a single sequence of frames taken by a camera action without any interruption.

  - Shot is regarded as a basic unit of constructing and analyzing a video content.

# Introduction(2/5)

- **Types of Shot Transitions:**

  Abrupt transition:

Cut

  Gradual transition:

Dissolve

Fade in

Fade out

Wipe

# Introduction(3/5)

- **What is a dissolve ?**

  -A dissolve sequence is the mixture of two video sequences, where the first sequence is fading out while the second one is fading in.

# Introduction(4/5)

- **Why reliable dissolve detection is still an unsolved problem ?**

  - Two gradually changing video sequences are not easy to be either temporally or spatially separated.

  -It is hard to distinguish dissolves from motions (global camera motion or local object motion).

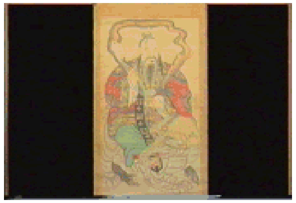
# Introduction(5/5)

- **The two shots before and after a dissolve effect may have**

(1)different color distributions,
different spatial layout.

(2)similar color distributions,
different spatial layout.

(3)similar color distributions,
similar spatial layout.

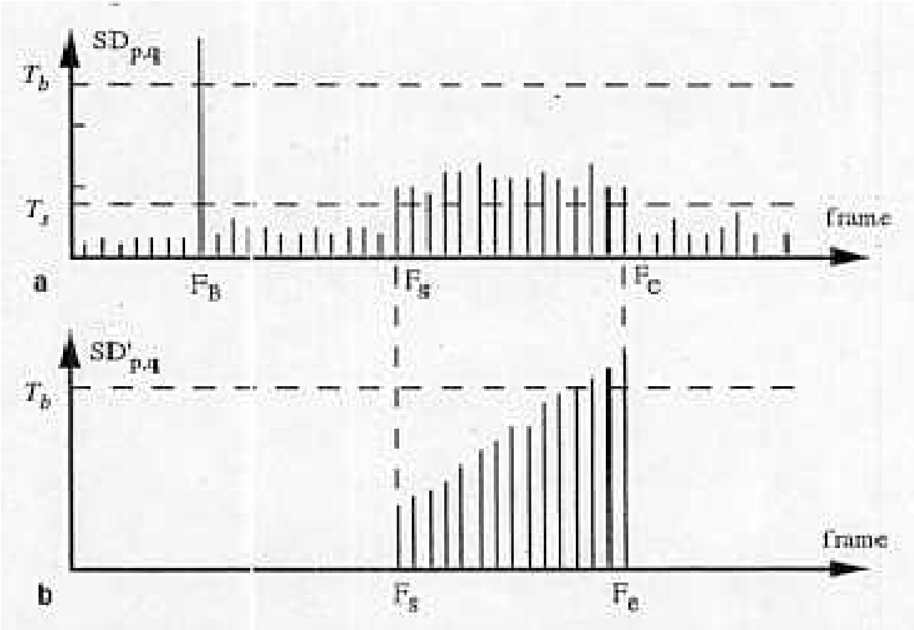


# The Categories of Shot Detection Methods

- Pixel-based.
- Histogram-Based.
- Statistic-based.
- Transform-based.
- Feature-based.
- Spatio-temporal Slice Model.

# Related Work(1/5) ~Histogram-based~

- Histogram-based(Zhang et al.'93):
  **Dual threshold values are applied to the difference of intensity histogram in order to detect gradual transitions.**



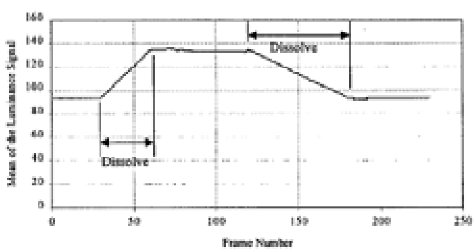

Twin comparison method

# Related Work(2/5)
# ~Statistic-based~

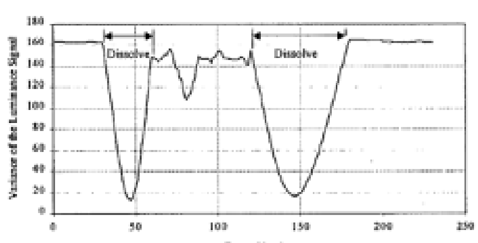
- Statistic-based(Fernando et al. ‘95):

**One can approximate that the mean and the variance have a linear and quadratic behavior during dissolving. Therefore,**

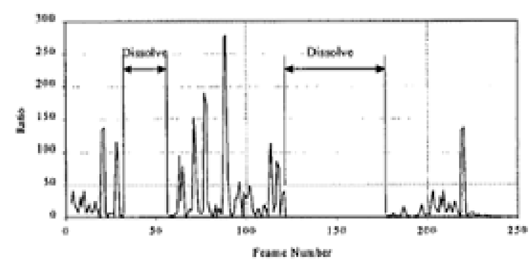
$$\frac{(\text{variance curve})''}{(\text{mean curve})'} = \text{Constant}$$




Mean of the luminance signal



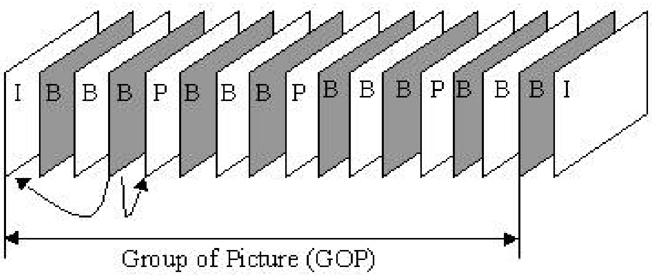

Variance of the luminance signal



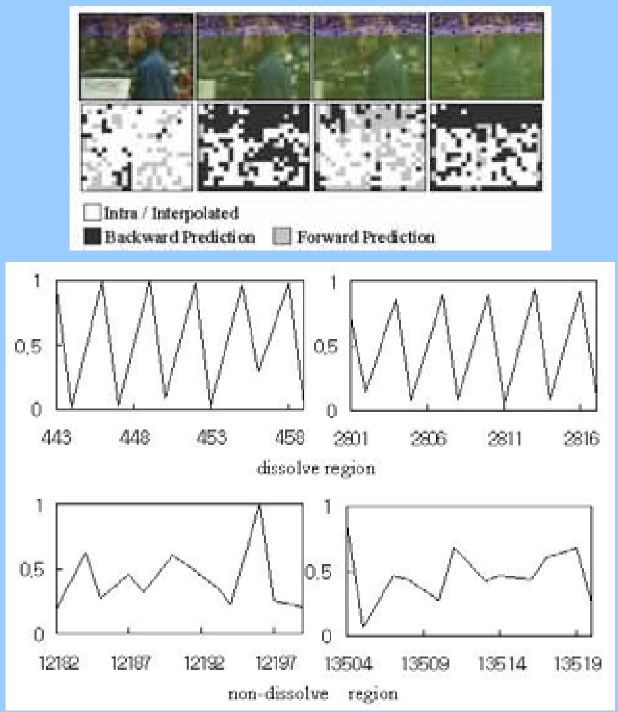

(variance curve)’’/(mean curve)’

# Related Work(3/5) ~Transform-based~

- Transform-based(Jun et al.‘00):

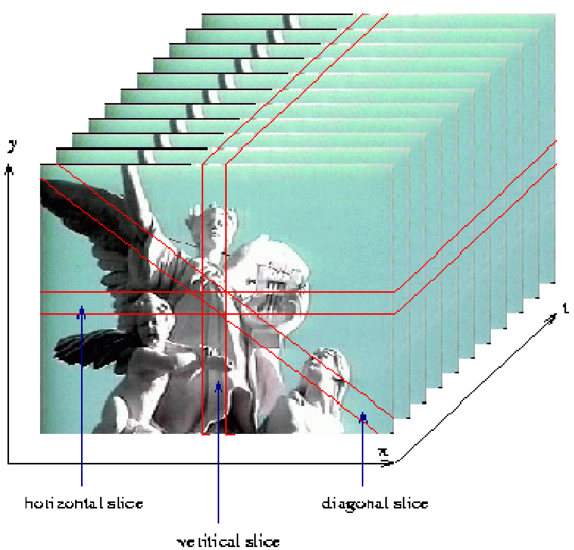


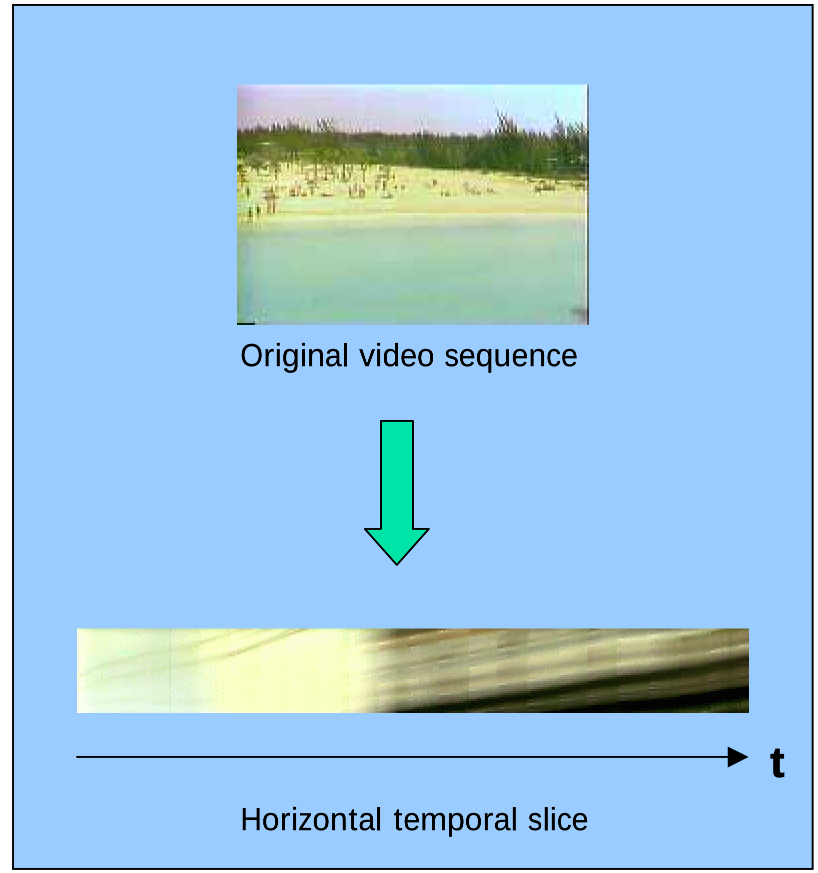

A typical MPEG video sequence




The ratio of forward macro block in the collected Sequence of B-type frames adjacent to anchor frames

# Related Work(4/5)
# ~Spatio-temporal Slice Model~

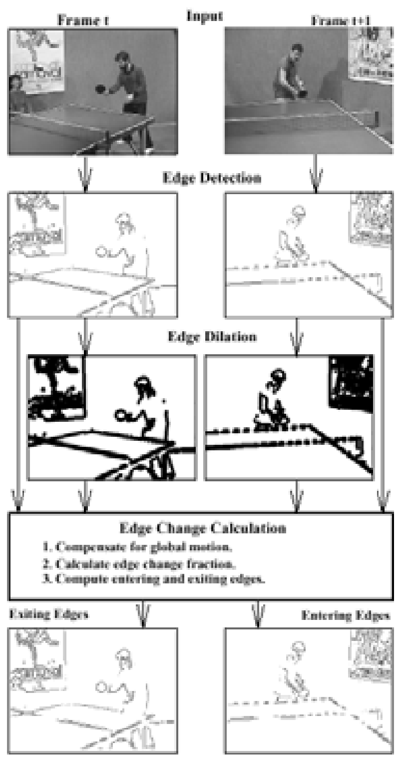


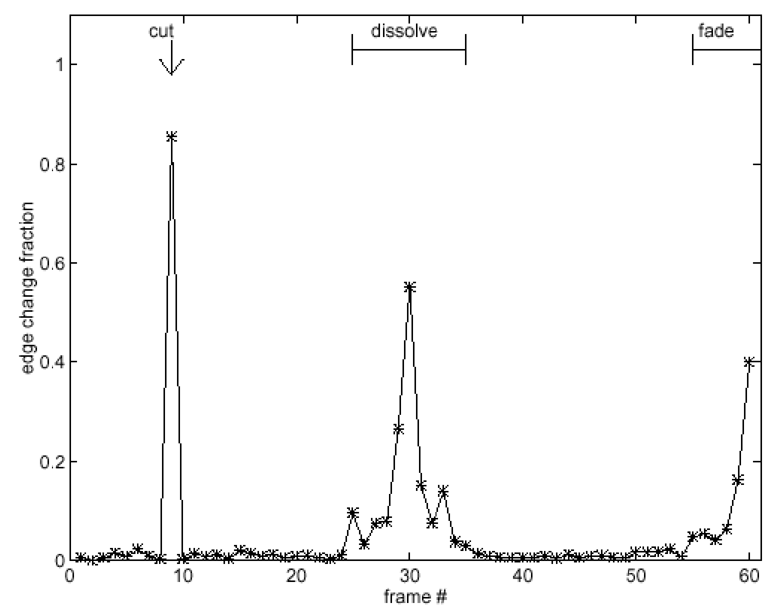

Images come from http://mmlib.cs.ust.hk/video.html

# Related Work(5/5) ~Feature-based~

- Feature-based(Zabih et al. ‘95):




Main steps of the computation of the edge change fraction




Results from the table tennis sequence

# Definition of a Dissolve(1/2)

- **An ideal n-frames dissolve can be expressed as follows:**

$$P_t(x,y)=A(t)F_t(x,y)+B(t)G_t(x,y), \quad \text{for } 0<t<n.$$

$P_t(x,y)$ is the pixel intensity positioned at (x,y) at the t-th frame.

$F_t(x,y)$ is the video sequence of the first shot.

$G_t(x,y)$ is the video sequence of the second shot.

# Definition of a Dissolve(2/2)

**Almost all methods start with two assumptions:**

**(1)Dissolve happens in strictly linear fashion.**

**(2)no or little motion in $F_t(x,y)$ & $G_t(x,y)$ sequences.**

**Hence, the transition can be further simplified as follows:**

$$P_t(x,y) = (1-t/n)F(x,y)+(t/n)G(x,y), \quad \text{for } 0<t<n.$$

$$\rightarrow P_{i+1}(x,y)-P_i(x,y) = (1/n)(F(x,y)-G(x,y))$$
$$= C(x,y), \text{ for } 0 \leqq i \leqq n-1.$$

# Dissolve Detection

## ～Our Idea ～

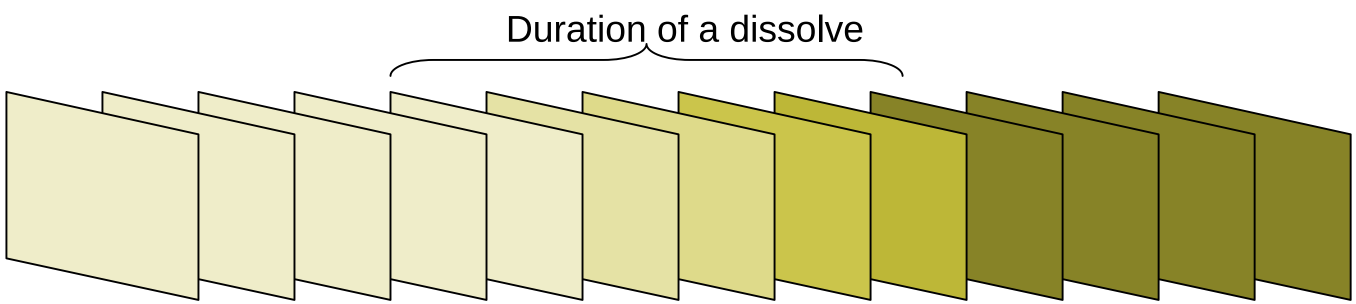
**Define $D^L_i(x,y)$ =** $$\begin{cases} 1, & if\ \ P_k(x,y) - P_{k-1}(x,y) > 0, \forall k \in [i-L+1, i] \\ 1, & if\ \ P_k(x,y) - P_{k-1}(x,y) < 0, \forall k \in [i-L+1, i] \\ 0, & otherwise \end{cases}$$

**, L = minimum number of frames in a dissolve**

→ **if** $\dfrac{\sum_{x,y} D_i^L(x,y)}{N} > \delta$ **,N = total number of pixels that changes in intensity value in a frame.**

→ **Dissolve may happen in [i-L+1, i].**

**e.g.:** If the dissolve contains 6 frames & $\sum D^6_9(x,y)/N = 1$ → Dissolve happens in [9-5, 9].




| Frame number | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 … |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\sum D^L_i(x,y)/N$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 … |

# Dissolve Detection

## ～Threshold(1/2)～

**Assume pixel intensity changes (increase or decrease) ,**

**P(increase) = P(decrease) = ½.**

**→ $P(D^L_i(x,y) = 1) = (½)^L \times 2 = (½)^{L-1}$.**

**→ $P(D^L_i(x,y) = 0) = 1 - P(D^L_i(x,y) = 1) = 1 - (½)^{L-1}$.**

**Suppose N pixels change in intensity due to some reasons, we can compute the cumulative density function, which is given by**

$$CDF(\delta) = \sum_{K=0}^{\delta} \binom{N}{K} \left(P\left(D_i^L(x,y)=1\right)\right)^K \left(P\left(D_i^L(x,y)=0\right)\right)^{N-K}$$

# Dissolve Detection

## ～Threshold(2/2)～

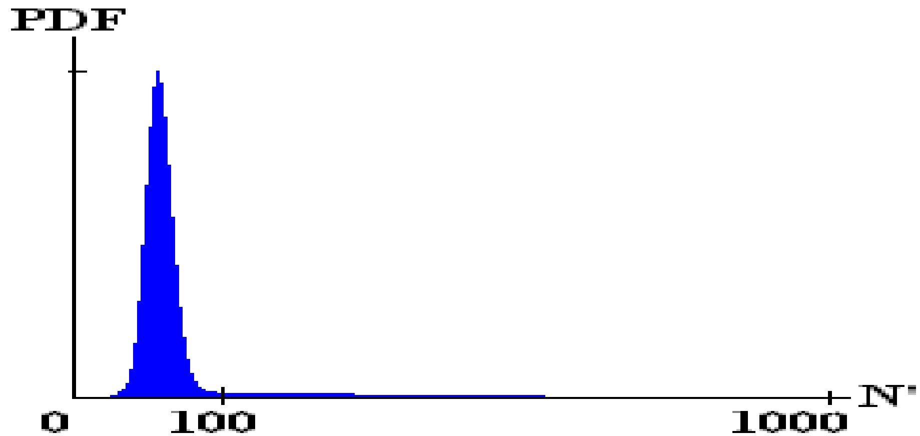
**The most common durations for dissolves are between 0.2 to 3 seconds, equivalent to 6 to 90 frames for 30 fps video sequences. Thus we decide L = 6-1 = 5.**

**→Suppose N = 1000, L = 5**

**→CDF(97) = 0.99999**

**→The probability of $\sum D^L_i(x,y)>97$ is Only 0.01%.**

**→We can define threshold = 0.1 > 97/1000**



# Dissolve Detection

## ～Problems～

- **Problem1: In real video sequences, object motion and camera motion may exist during dissolves. And pixel intensity also slightly change their values in a compressed video.**

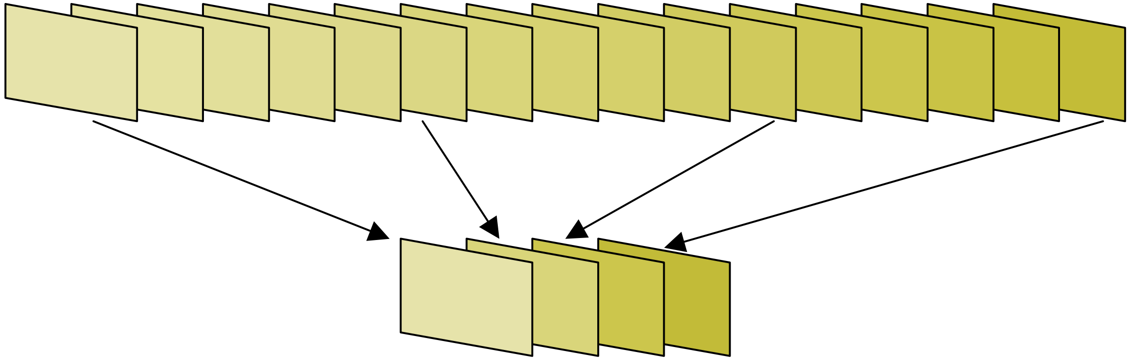
- **Problem2: $|P_{i+1}(x,y) - P_i(x,y)| \approx 0$ when the duration of a dissolve is long, $P_{i+1}(x,y) - P_i(x,y)$ may not be continuous by increasing or decreasing.**

- **Problem3: false detection may be caused by having a light source involved, e.g. sunlight, lamps, or an explosion.**

# Dissolve Detection

## ～improvement～

- **In order to solve the second problem, we subsample the frames of the original video sequence.**



- **In order to solve the third problem, If pixel intensity increase for all pixels in a detected dissolve, and the Last frame is a lighten frame. Then the dissolve may be a false detection due to a light source.**

# Conclusion

- Advantages of our method

1. Even a dissolve with the mixture of two similar (color or spatial) video sequences can be detected by our method.

2. Clearly distinguish dissolves from motions.

3. Dissolves can be detected, no matter how long a duration is.

# Future Work

- **It's still a difficult problem to detect dissolves containing significant motion. We can change the pixel-based method into block-based method, or directly take advantage of the DC value of an 8 x 8 luminance block in an MPEG stream.**