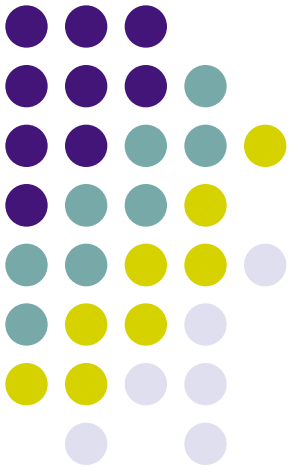
Camera Attention Weighted Strategy for Video Shot Grouping

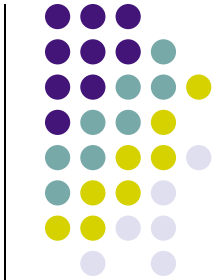
Fan Jiang, Yu-Jin Zhang

Electronic Engineering Dept.
Tsinghua University

2005.4

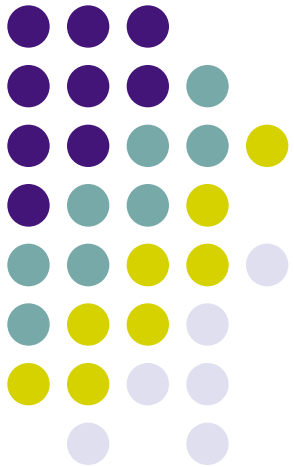


Outline



- Introduction
- Shot Feature Extraction
- Shot Grouping Scheme
- Experimental Results
- Conclusion

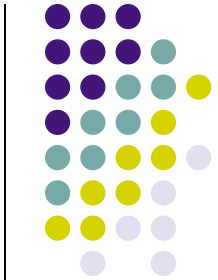
Introduction



Introduction (1)

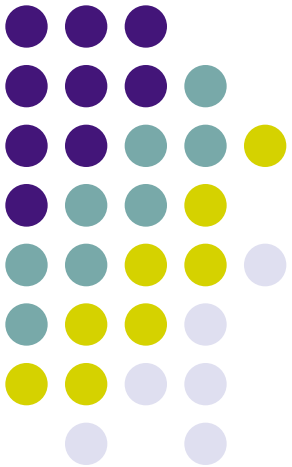
- Shot grouping
  - Video structuring above shot-level

Introduction (2)



- Visual feature extraction
  - Frame-based approach
    - Global: no semantics
  - Object-based approach
    - Crude and unreliable
  - Temporal Weighting
    - Uneven importance given to different time

Shot Feature Extraction



Shot Feature Extraction (1)

- Camera motion analysis
  - Four-parameter global motion model

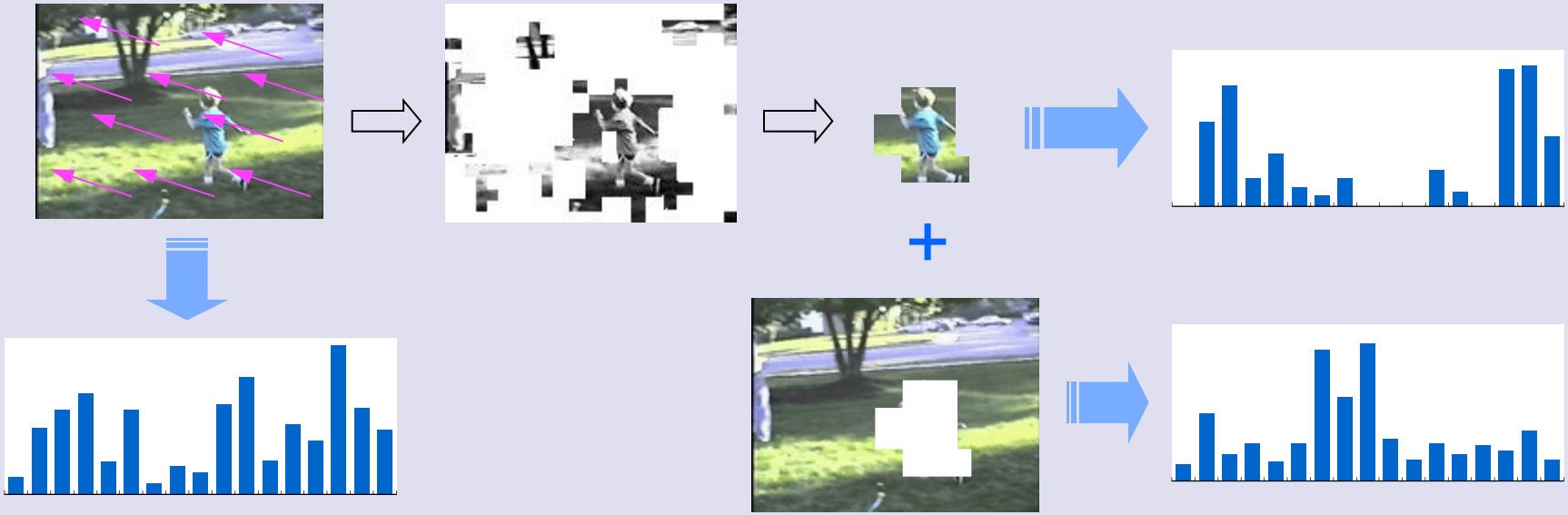
$$\overline{MV} = \begin{bmatrix} a_1 & a_2 \\ -a_2 & a_1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} a_3 \\ a_4 \end{bmatrix}$$

  - Motion vector pre-processing

$$\begin{cases} \overline{FlowI} = \dfrac{1}{2}(\overline{MV}_{preB,forward} - \overline{MV}_{nextB,backward}) \\ \overline{FlowB} = \dfrac{1}{2}(\overline{MV}_{forward} - \overline{MV}_{backward}) \\ \overline{FlowP} = \dfrac{\overline{MV}_{forward}}{Nref} \end{cases}$$

Shot Feature Extraction (2)

- Frame segmentation
  - Attention Region (AR) + Background (BG)



Shot Feature Extraction (3)

- Camera attention weighted strategy
  - Real camera moving parameter

$$\overline{MV} = \begin{bmatrix} a_1 & a_2 \\ -a_2 & a_1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} a_3 \\ a_4 \end{bmatrix}$$

$$\begin{cases} s = a_1 + 1 \\ \gamma = a_2 / (a_1 + 1) \\ L = \sqrt{(f\alpha)^2 + (f\beta)^2} = \sqrt{{a_3}^2 + {a_4}^2} \Big/ (a_1 + 1) \\ \theta = arctg \dfrac{f\beta}{f\alpha} = -arctg \dfrac{a_4}{a_3} \end{cases}$$

Shot Feature Extraction (4)

- Camera attention model
  - Frames with attention regions

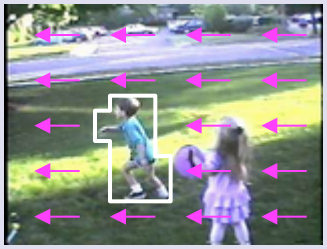
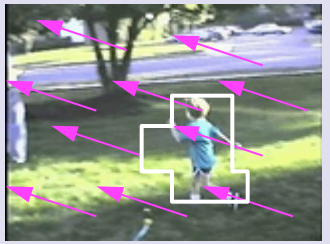
$$wtBG = \frac{1}{wtAR} \qquad wtAR = \begin{cases} s & L < L_0 \\ s \cdot (1 + L / r_L) & L \geq L_0 \end{cases}$$

  - Frames without attention regions

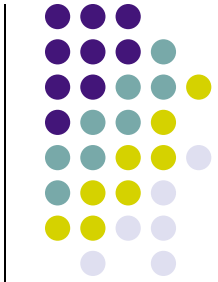
$$wtBG = \begin{cases} s & L < L_0 \\ s/(1 + f(\theta) \cdot L / r_L) & L \geq L_0,\ \theta < \pi/4 \\ s \cdot (1 + f(\theta) \cdot L / r_L) & L \geq L_0,\ \theta \geq \pi/4 \end{cases}$$

Shot Feature Extraction (5)

- Examples of camera attention weights
  - Camera Pan:

			
wtBG	0.924	0.856	0.583
wtAR	/	1.168	1.715

Shot Feature Extraction (6)



- Weighted histogram

$$\begin{cases} HistAR_i = \sum_t histAR_i(t) \cdot wtAR_i(t) \quad (i = 1, 2, \ldots, M) \\ HistBG = \sum_t histBG(t) \cdot wtBG(t) \end{cases}$$

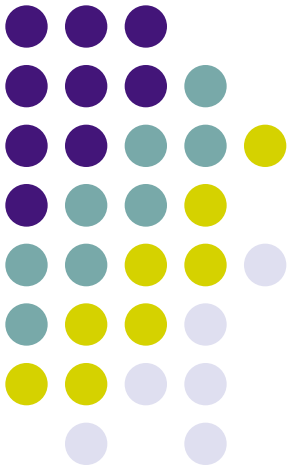
- Attention regions weights

$$\left\{ WtAR_i = \frac{\sum_t wtAR_i(t)}{\sum_i^M \sum_t wtAR_i(t)} \right\}_{i=1}^{M}$$

- Shot feature vector

$$F = \left\{ HistBG, \ \left\{ HistAR_i, WtAR_i \right\}_{i=1}^{M} \right\}$$

Shot Grouping Scheme

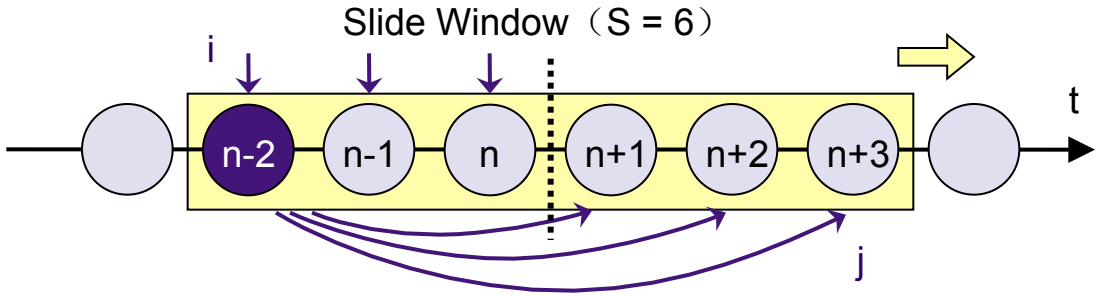


Shot Grouping Scheme (1)

- Shot similarity measures

$$\begin{cases} SimBG_{mn} = \sum_{k} Min(HistBG_m(k), HistBG_n(k)) \\ SimAR_{mn} = \sum_{i,j} \sum_{k} Min(HistAR_{m,i}(k), HistAR_{n,j}(k)) \cdot WtAR_{m,j} \cdot WtAR_{n,j} \end{cases}$$

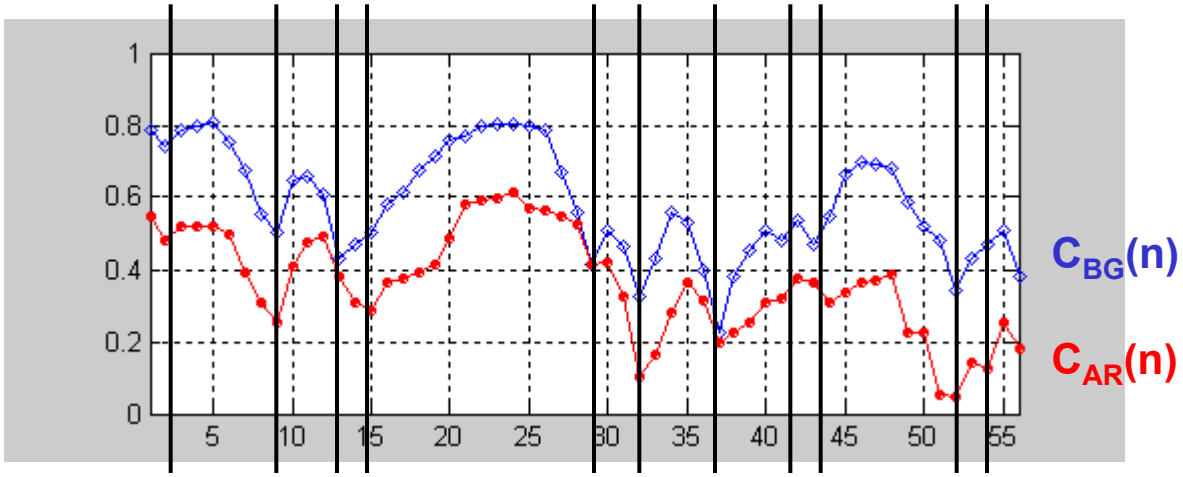
- Sequential grouping approach



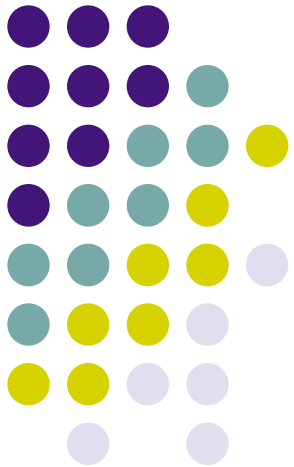
  - C(n) = ave $\{Sim_{ij}\}_{sld\_wnd}$

Shot Grouping Scheme (2)

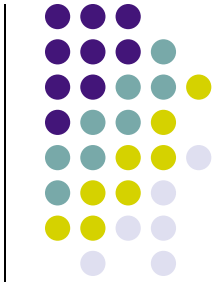
- Shot groups segmentation
  - Local minimum of C(n)



Experimental Results



Experimental Results



- Shot grouping results
  - 3 clips / 868 shots / 170 scenes

Similarity measure	Global Feature Measure			Our Measure		
Hits	5	10	109	5	12	122
Misses	0	5	41	0	3	28
False alarms	3	1	12	1	1	19

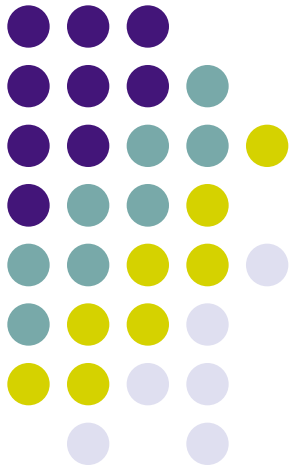
Recall = 72.9%

Precision = 88.6%

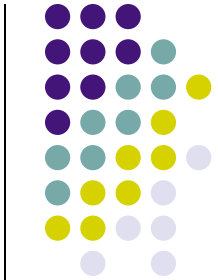
Recall = 81.8%

Precision = 86.9%

Conclusion



Conclusion



- Attention-based feature extraction
  - Spatial segmentation: AR + BG
  - Temporal weighting: by camera motion
- Shot grouping based on two similarities
- Operating on MPEG domain
- Further direction
  - Multi-AR
  - Multi-level shot organization

Thanks