SIGGRAPH ASIA 2017 BANGKOK

CONFERENCE 27 – 30 November 2017
EXHIBITION 28 – 30 November 2017
BITEC, Bangkok, Thailand
SA2017.SIGGRAPH.ORG

Sponsored by acm

Multi-video Object Synopsis Integrating Optimal View Switching

Zhensong Zhang[1], Yongwei Nie[2], Hanqiu Sun[1], Qiuxia Lai[1], Guiqing Li[2]

[1]The Chinese University of Hong Kong
[2]South China University of Technology

INTRODUCTION

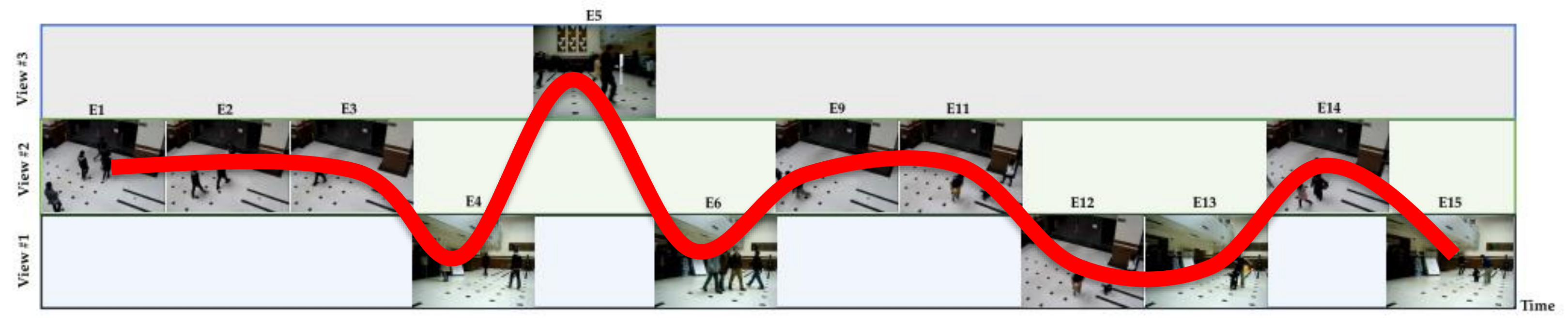
- Surveillance cameras are widely installed
  - Contain a lot of redundant information
  - Challenges to storage and browsing
  - Various methods, e.g., video fast forward, video skimming, video abstraction, video summarization, video synopsis

INTRODUCTION

- We study the synopsis problem of multi-view surveillance videos
  - Public areas (e.g., bank, museum, hospital, et al.)
  - With overlapped field of view
  - A small set of cameras can generate large data

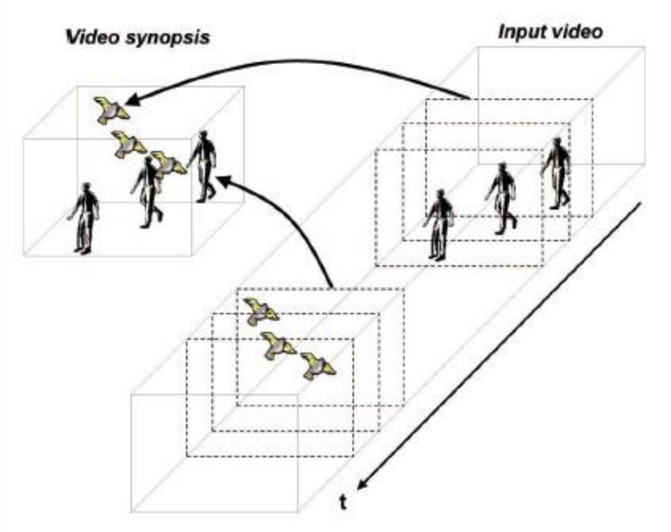
INTRODUCTION

- Frame-based approaches [Fu et al. 2010; Panda et al. 2017]
  - Connect key frames, one video, view switching
  - Discrete views of moving objects
  - Retain redundancies in regions with no activity



INTRODUCTION

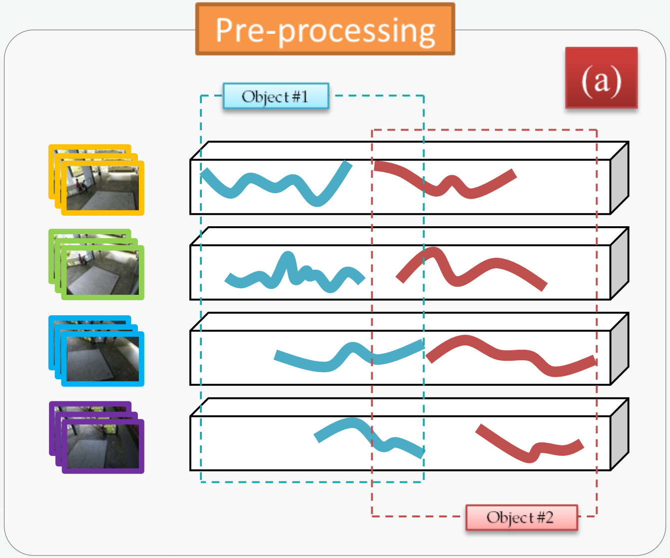
- Object-based approaches [Pritch et al. 2008; Zhu et al. 2016; Mahapatra et al. 2016]
  - Extract path of moving object as tube
  - Directly shift object tubes in spatiotemporal space
  - Display the synopsized videos in a side by side manner



OUR METHOD

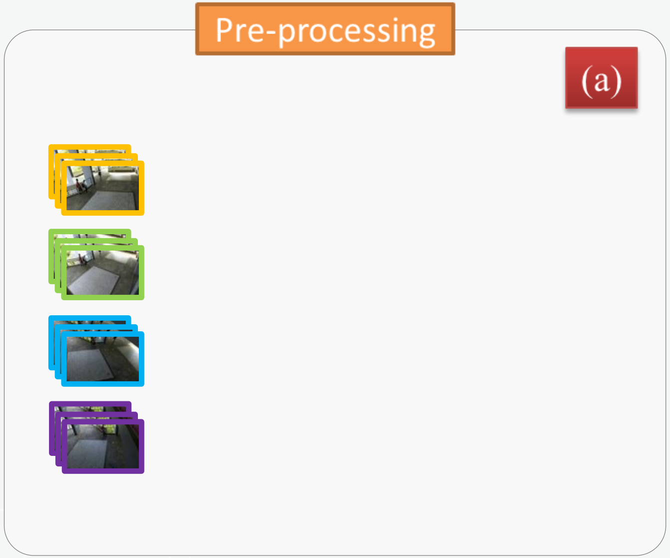
- We propose the unified framework for multi-video object-based synopsis
  - Utilize the advantages of both approaches
    - Condense videos by shifting object tubes in temporal space
    - Display unified synopsis by optimal view switching
  - Joint optimization
    - Optimal temporal offsets of moving objects
    - Optimal viewing sequence

OUR METHOD



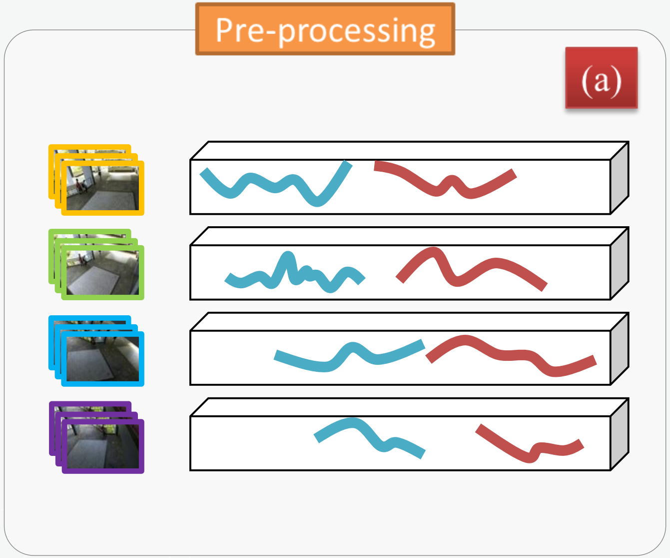
- Preprocessing
  - Synchronize all the input videos
  - Extract the static background of each video
  - Detect all the objects in the videos as spatiotemporal tubes
  - Group the tubes of the same object in different videos as a single object

OUR METHOD



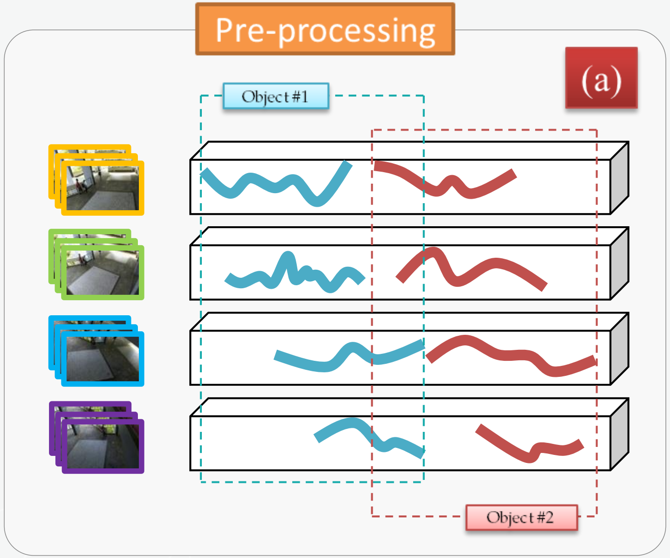
- Preprocessing
  - Synchronize all the input videos
  - Extract the static background of each video
  - Detect all the objects in the videos as spatiotemporal tubes
  - Group the tubes of the same object in different videos as a single object

OUR METHOD



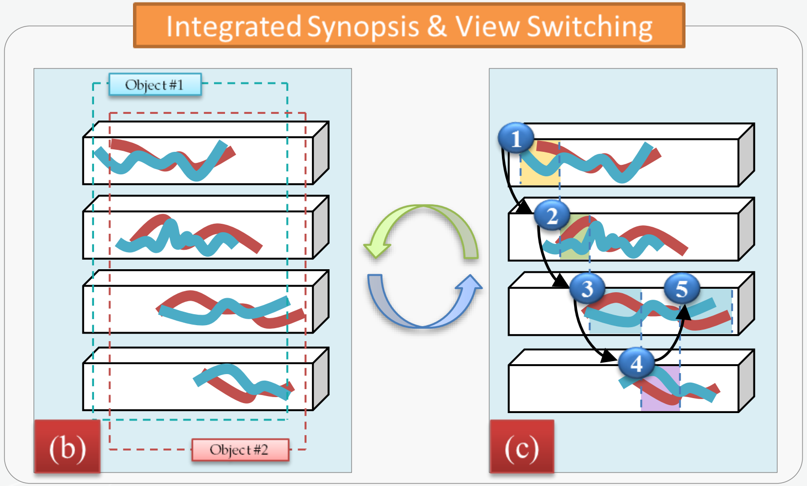
- Preprocessing
  - Synchronize all the input videos
  - Extract the static background of each video
  - Detect all the objects in the videos as spatiotemporal tubes
  - Group the tubes of the same object in different videos as a single object

OUR METHOD



- Preprocessing
  - Synchronize all the input videos
  - Extract the static background of each video
  - Detect all the objects in the videos as spatiotemporal tubes
  - Group the tubes of the same object in different videos as a single object

OUR METHOD



- **The temporal offset** for moving objects
  - $T = \{\delta_i | i \in [1, N]\}$
- **The optimal view** we select for each frame in the synopsis
  - $K = \{k_t | k_t \in [1, M], t \in [1, L]\}$

OUR METHOD

- Our unified framework for multi-video synopsis

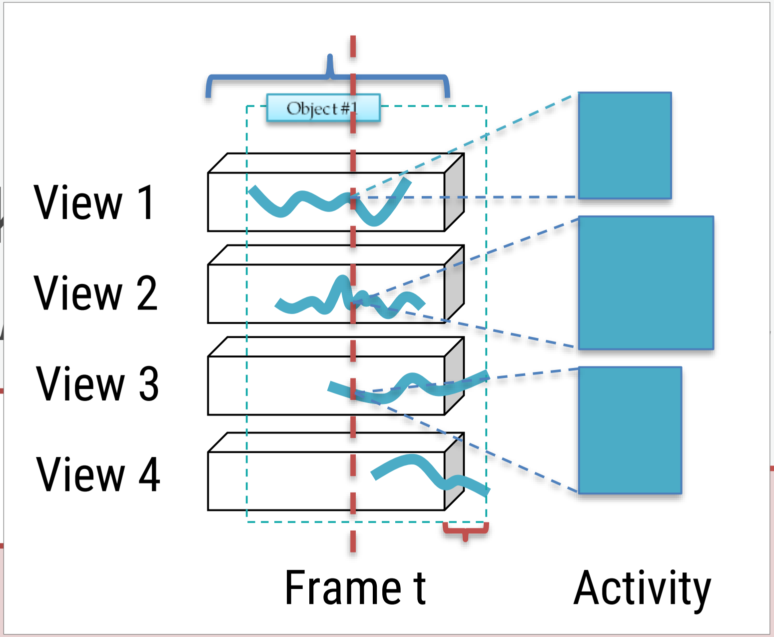
$$E(T,K) = w_a E_a(T,K) + w_o E_o(T,K) + w_{co} E_{co}(T) + w_{vs} E_{vs}(K)$$

Activity cost	Occlusion cost	Chronological order cost	View switching cost
Preserve activities	Avoid occlusions	Keep chronological orders of objects	Switch between views optimally

ACTIVITY COST

- Our unified framework

$$E(T,K) = w_a E_a(T,K) + w_o \ldots (K)$$

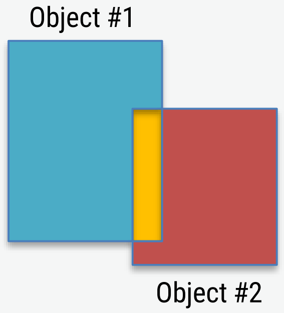


Activity cost E_a

- Preserves activities
- Composed of two parts
  - Penalizes all the activities that are not included in the synopsized videos
  - Favors the view that contains more activities

SIGGRAPH ASIA 2017 BANGKOK

OCCLUSION COST



- Our unified framework

$$E(T,K) = w_a E_a(T,K) + w_o E_o(T,K) + w_{co} E_{co}(T) + w_{vs} E_{vs}(K)$$

Occlusion cost E_o

- Prevents collisions between shifted objects
- Sum of the overlapping ratio between all pairs of objects over all selected views

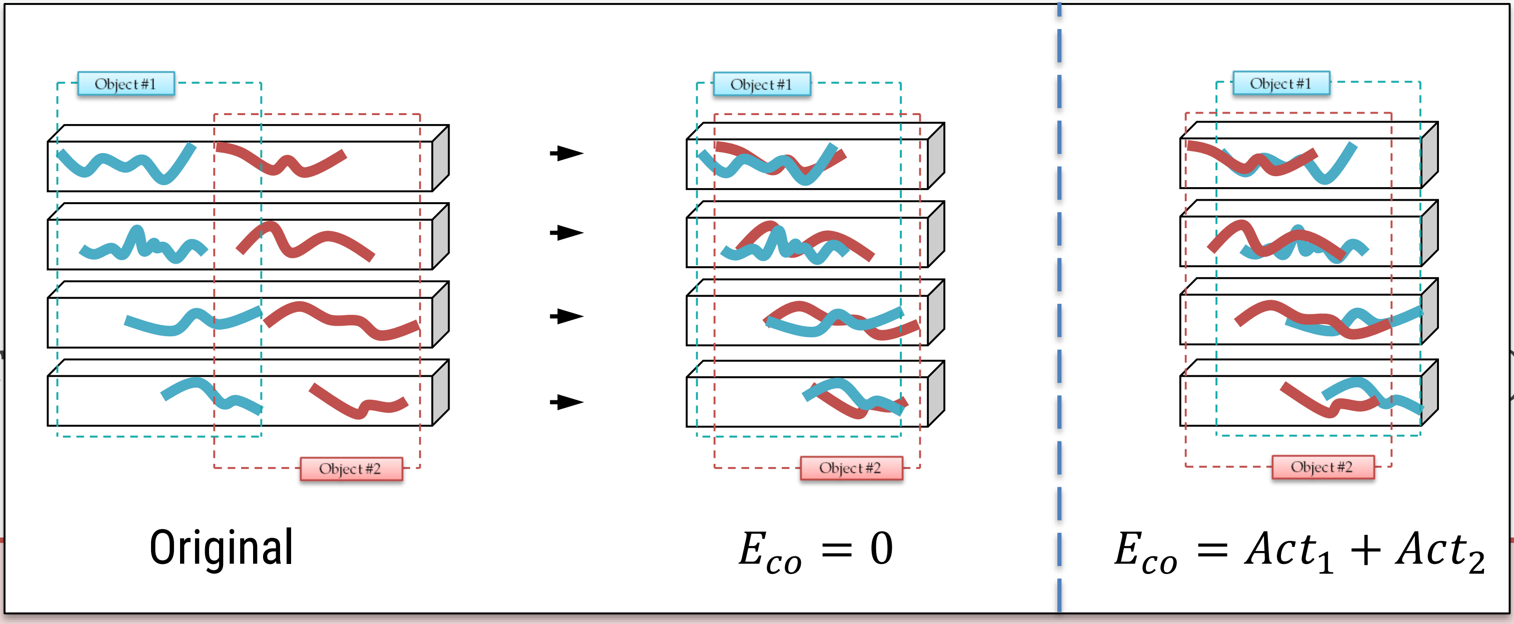
CHRONOLOGICAL ORDER COST

- Our unified framework

$$E(T,K) = w_a E_a(T,K) + w_o E_o(T,K) + w_{co} E_{co}(T) + w_{vs} E_{vs}(K)$$

Chronological order cost E_{co}

- The objects appear first in the input videos should also appear first in the synopsis video
- Sum of activities violating the chronological order



- The objects appear first in the input videos should also appear first in the synopsis video
- Sum of activities violating the chronological order

VIEW SWITCHING COST

- Our unified framework

$$E(T,K) = w_a E_a(T,K) + w_o E_o(T,K) + w_{co} E_{co}(T) + w_{vs} E_{vs}(K)$$

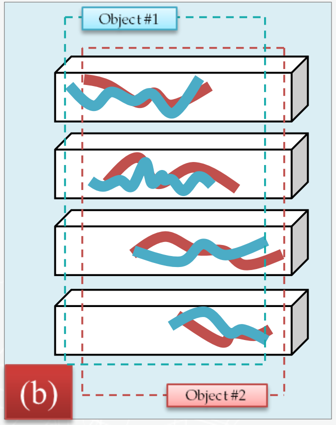
View switching cost E_{vs}

- Ensures the viewpoints to be changed optimally
- Constrain the current view to be the same as the last view
- Constrain a view should be continuously selected at least 30 frames and at most 150 frames

OPTIMIZATION

- Our unified framework

$$E(T,\cancel{K}) = w_a E_a(T,\cancel{K}) + w_o E_o(T,\cancel{K}) + w_{co} E_{co}(T) \cancel{+ w_{vs} E_{vs}(K)}$$

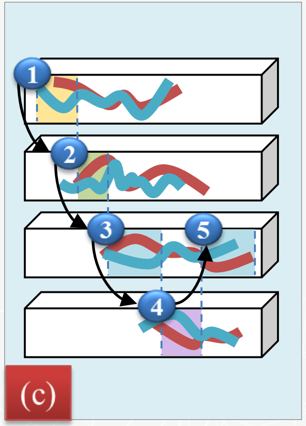


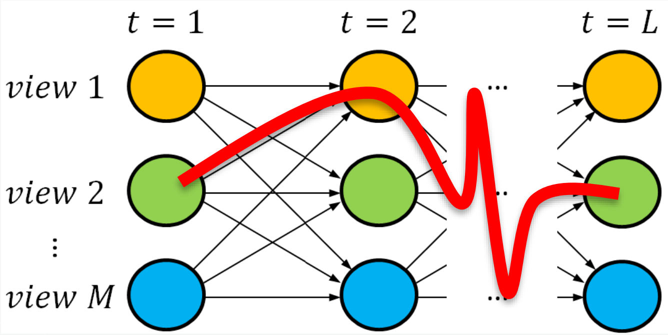
- Step 1: optimize T first by fixing K
- Reduced to the sum of energy terms defined on single object (E_a) or pair of objects (E_o, E_{co})
- Graph Cuts

OPTIMIZATION

- Our unified framework

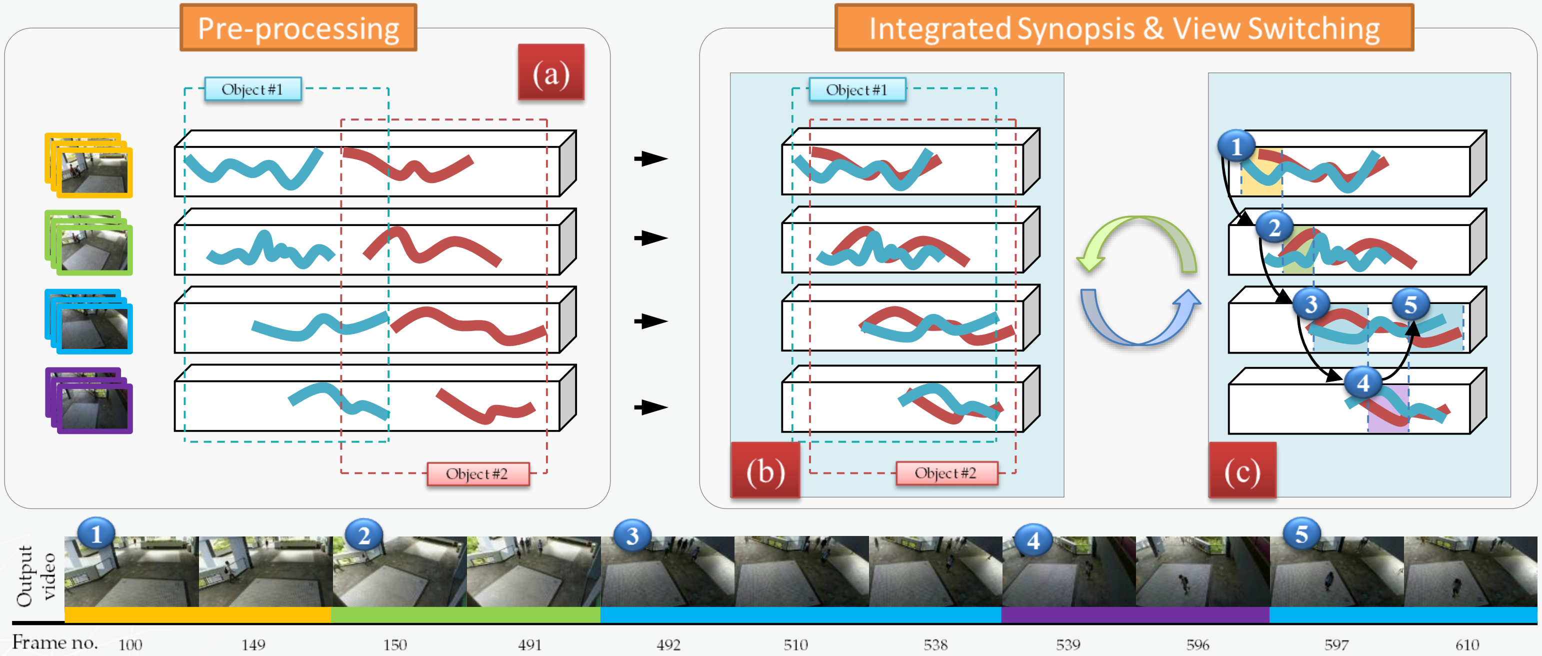
$$E(\cancel{T,}K) = w_a E_a(\cancel{T,}K) + w_o E_o(\cancel{T,}K) \cancel{+ w_{co} E_{co}(T)} + w_{vs} E_{vs}(K)$$





- Step 2: optimize K with the known T
- Node weight: $E_a + E_o$
- Edge weight: E_{vs}
- Dynamic programming [Arev et al. 2014]

OUR METHOD: SUMMARY



EXPERIMENTS

	Case 1	Case 2
Video Number	2	4
Video Length	4 min	10 min
Frame Rate	30 fps	30 fps
Angle	> 90°	About 20°
Synopsis Length	33 sec	33 sec
Video Comparison		Frame-based approach

DISCUSSION

Comparisons with previous object-based methods

Method	Feature	Overlapped Fov	# of Output Videos
[Zhu et al. 2016]	Multicamera joint video synopsis, consider the chronological orders of tubes from different views	No	Multiple
[Hoshen and Peleg 2015]	One live Master camera, one or more Slave cameras showing past activities	No	Multiple
[Mahapatra et al. 2016]	Mapping multiple views to a common ground plane	Yes	One
Our Method	Display multi-video synopsis by optimal view switching	Yes	One

EXPERIMENTS

- Performance
  - 10 iterations at most
  - Graph Cuts: 1 – 10 seconds
  - Dynamic Programming: < 0.1 second

- Time marks
  - Show moving objects with the *original appearance timestamps*

EXPERIMENTS

Multi-video Object Synopsis Integrating Optimal View Switching

Zhensong Zhang[1] Yongwei Nie[2] Hanqiu Sun[1] Qiuxia Lai[1] Guiqing Li[2]

[1]The Chinese University of Hong Kong
[2]South China University of Technology

SUMMARY

- A unified framework for multi-video object-based synopsis
  - The occurrences of moving objects in different cameras are treated identically
  - Jointly optimize temporal offsets and optimal view sequence
  - Unified synopsis space of multiple input videos
- Further work
  - User interaction preferences (i.e., frame #, zoom view)
  - More constraints for view selection

CONFERENCE 27 – 30 November 2017
EXHIBITION 28 – 30 November 2017
BITEC, Bangkok, Thailand
SA2017.SIGGRAPH.ORG

Sponsored by

THANK YOU!

Zhensong Zhang, Hanqiu Sun, Qiuxia Lai

-{zszhang, hanqiu, qxlai}@cse.cuhk.edu.hk

Yongwei Nie, Guiqing Li

-{nieyongwei, ligq}@scut.edu.cn