

NON-CHRONOLOGICAL VIDEO SYNOPSIS AND INDEXING

Presented by *Mikel Rodriguez*

Yael Pritch Alex Rav-Acha Shmuel Peleg

Problem Introduction

- Millions of surveillance cameras around the world
 - Constantly recording video
- Time consuming to review each one



Objective

- Generate a video summary

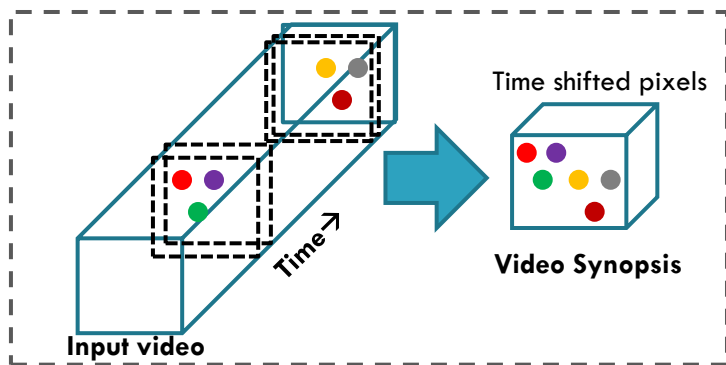
Input: Long video



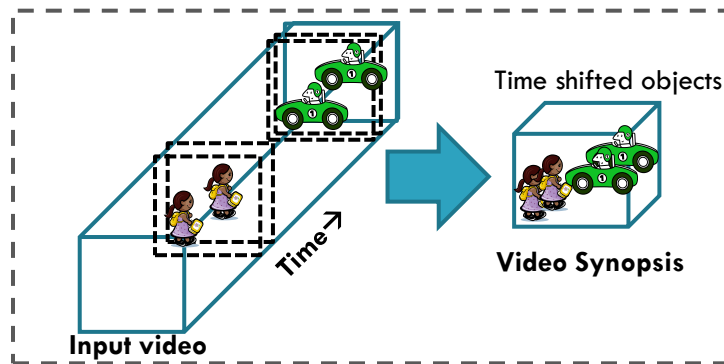
Output: Short video

Approaches Overview

Low-level video summary



Object-based video summary



A horizontal bar at the top of the slide, divided into a red section on the left and a teal section on the right. The text 'Low-level video summary' is written in white on the teal background.

Low-level video summary

PIXEL-BASED

Synopsis by energy minimization

- Goal: Find a mapping (M) for each pixel that minimizes:

$$E(M) = E_a(M) + \alpha E_d(M)$$

Activity cost

Discontinuity cost

- The mapping M represents time shifts

Activity Cost

- Number of foreground pixels not in summary video
- Foreground pixel:

$$\chi(x, y, t) = \| I(x, y, t) - B(x, y, t) \|$$

Respective pixel in
background image

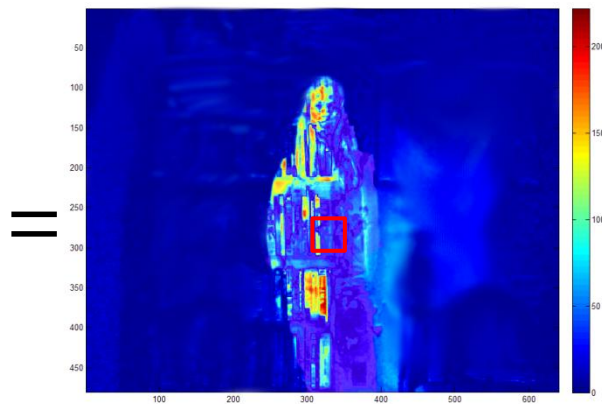
Pixel in the input image



-

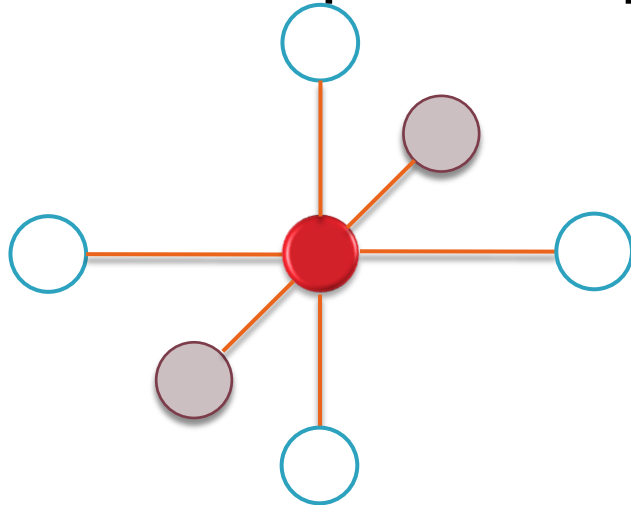


=



Discontinuity Cost

- Does a pixel shift preserve its neighborhood ?
- Look at a spatio-temporal neighborhood:



 4 spatial neighbors

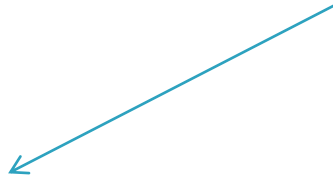
 2 temporal neighbors

Discontinuity Cost

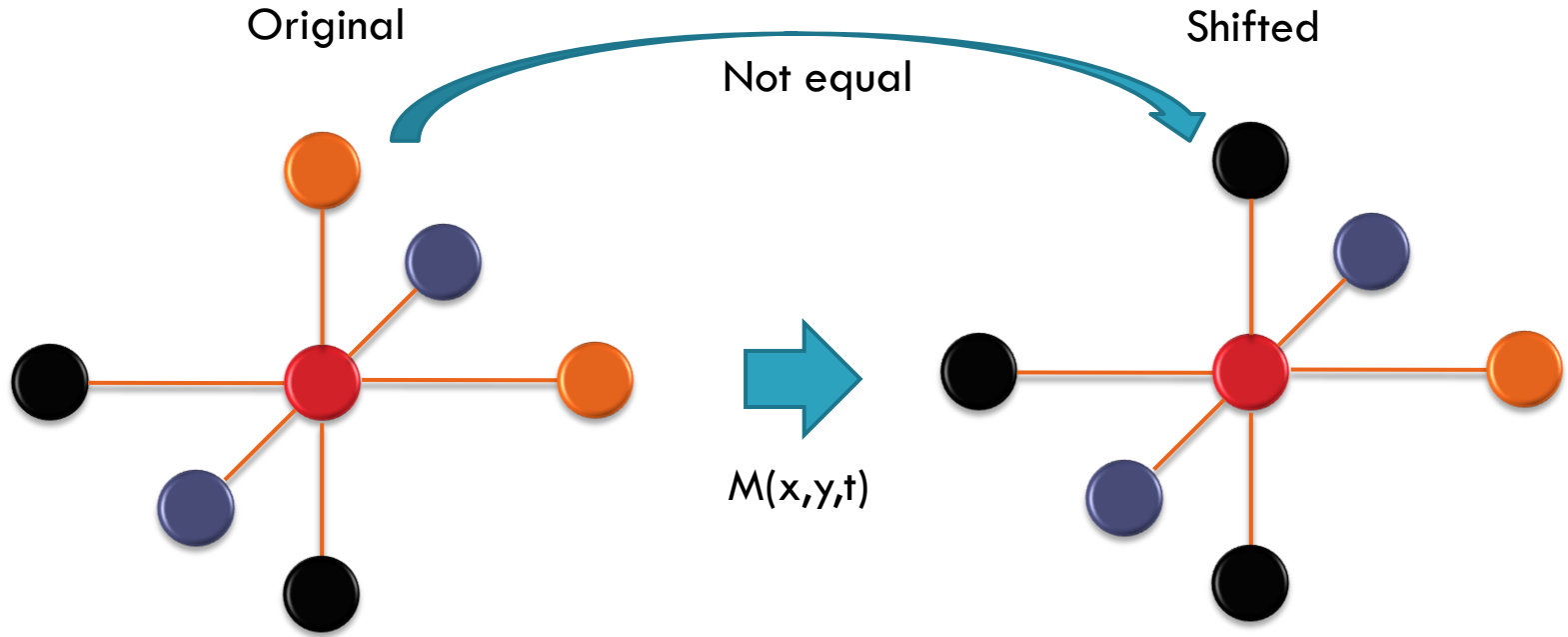
Original



A pixel in the input video

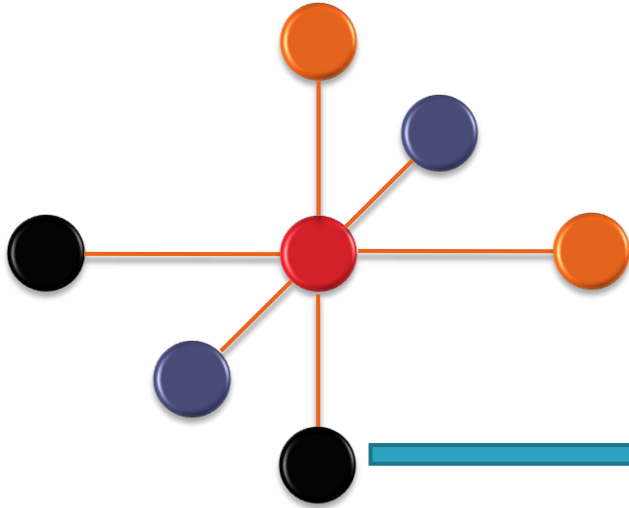


Discontinuity Cost

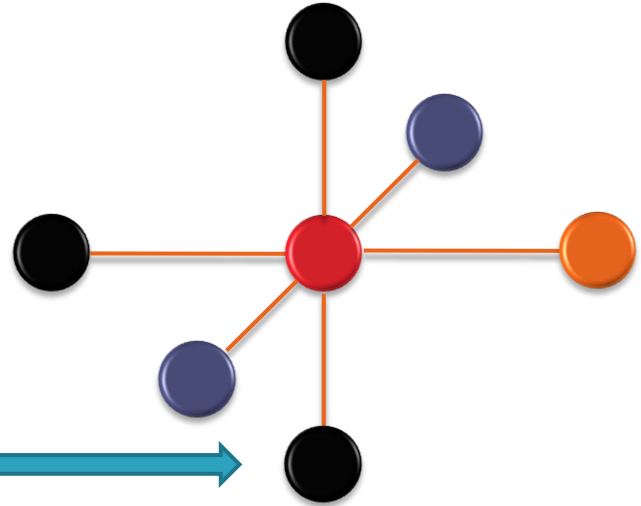


Discontinuity Cost

Original

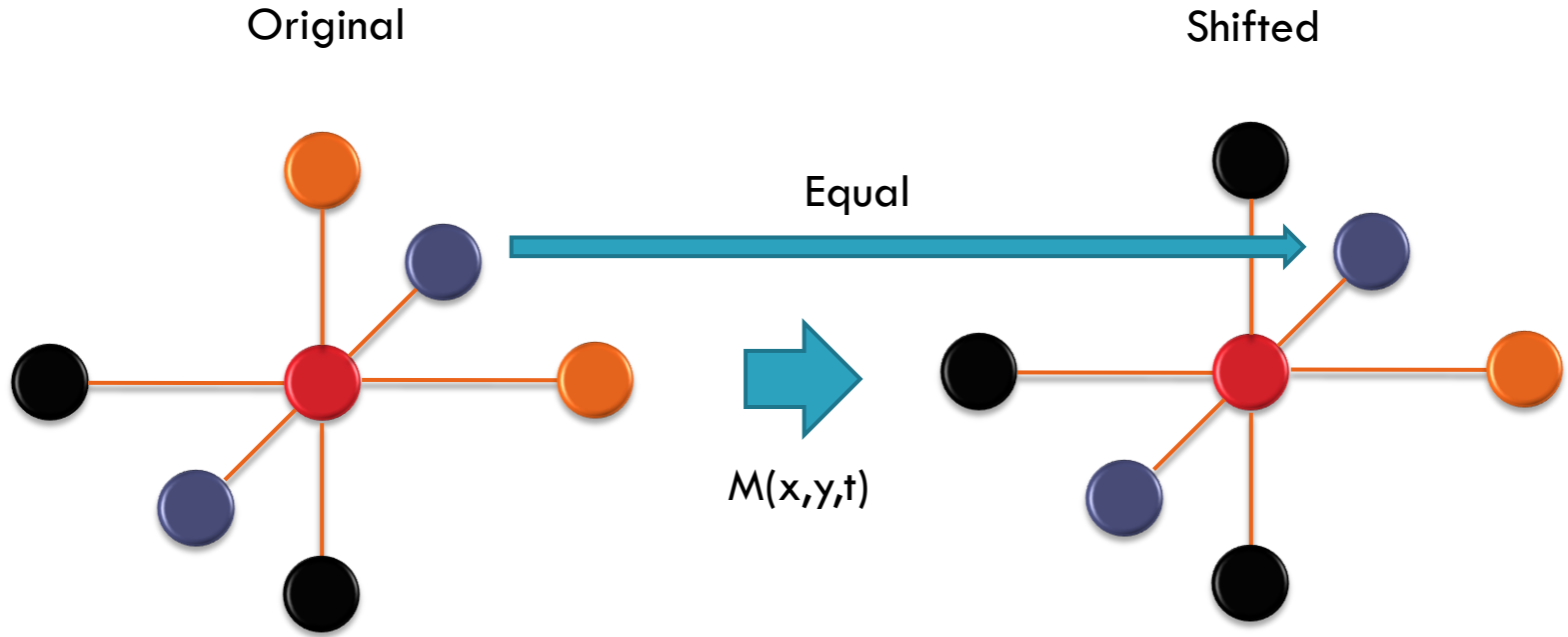


Shifted

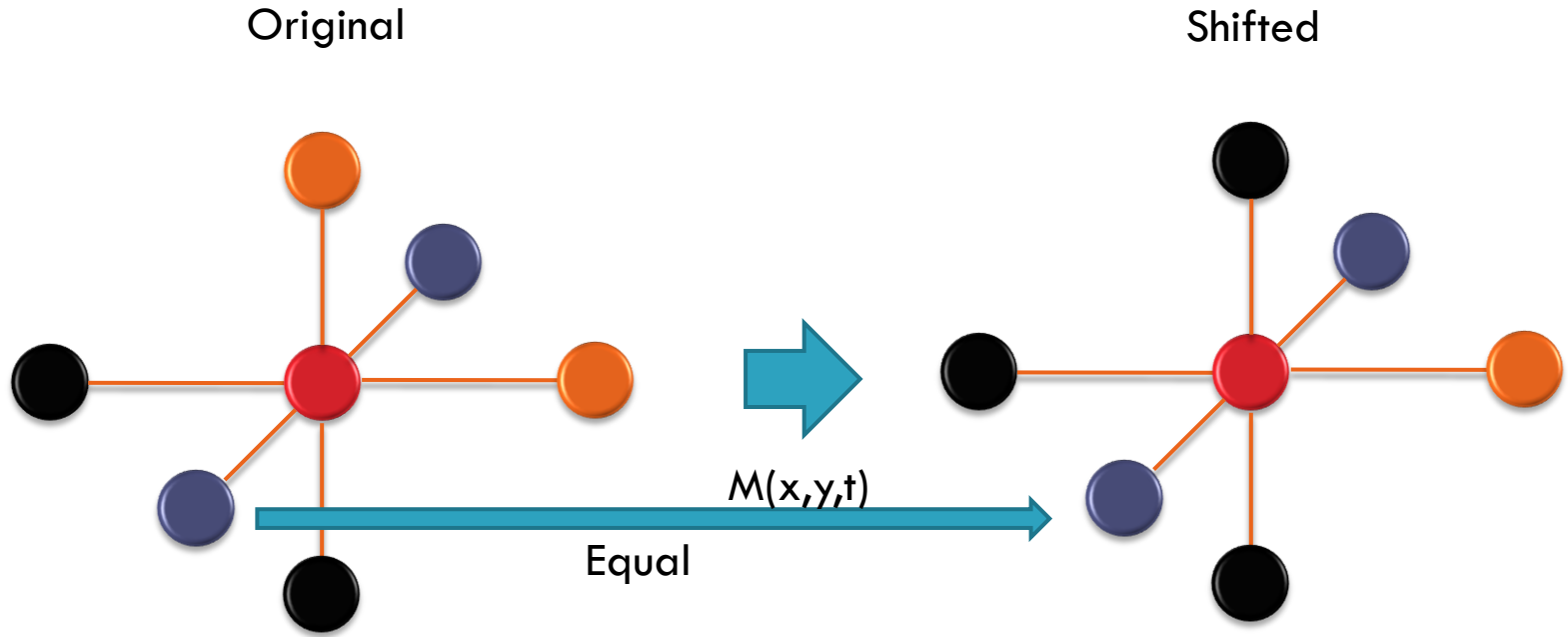


Equal

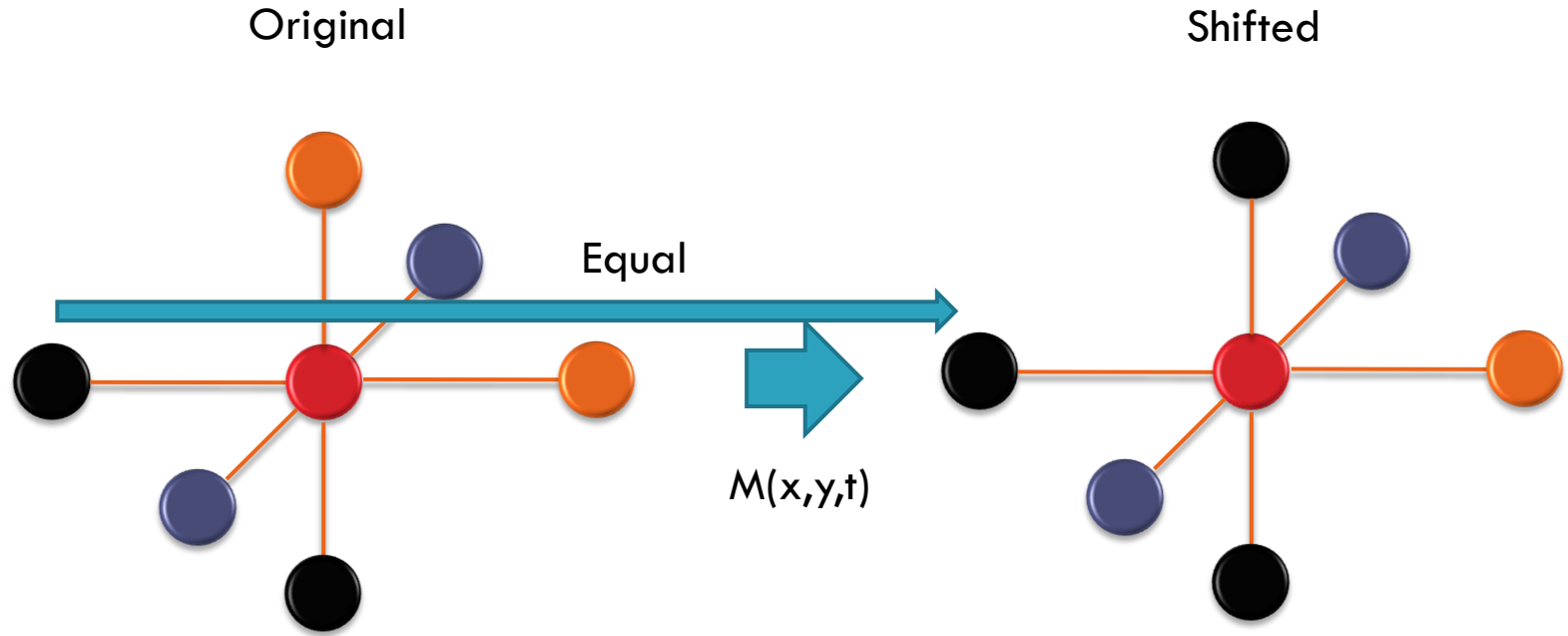
Discontinuity Cost



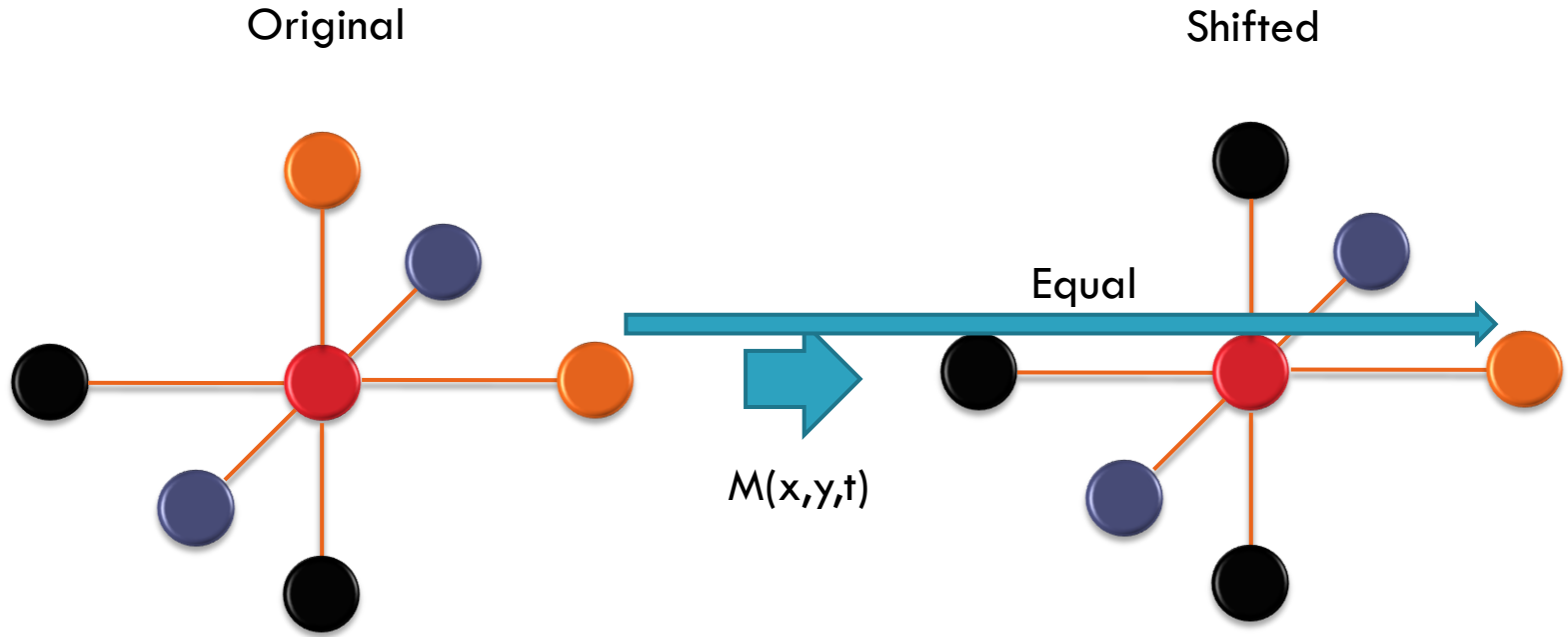
Discontinuity Cost



Discontinuity Cost



Discontinuity Cost



Drawbacks of Pixel-based Summary

- Limited to satisfying local properties:
 - ▣ Such as visible seams.
- High computational cost
- Solution:
 - ▣ Incorporate high-level analysis

A horizontal bar at the top of the slide, divided into a red section on the left and a teal section on the right. The text "Object-based Video Synopsis" is written in white on the teal section.

Object-based Video Synopsis

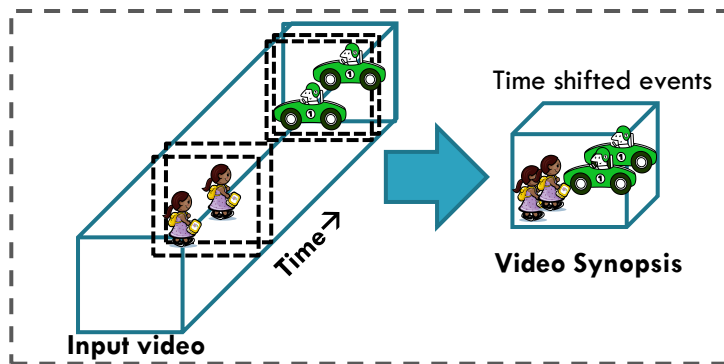
Object-based Video Synopsis

Objective:

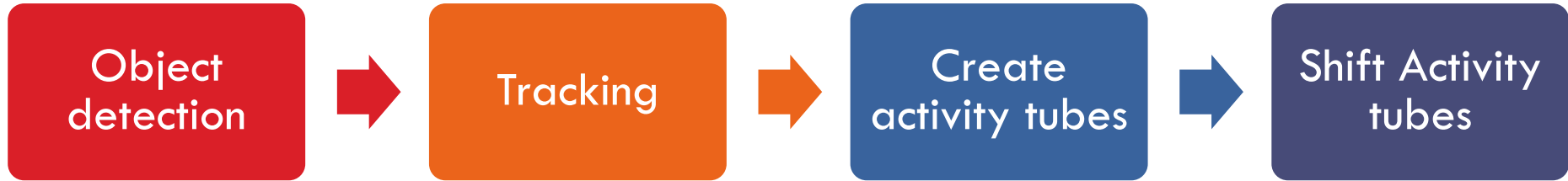
- Create synopsis of a video
- Combine activities from different times

Approach:

- Object detection and tracking.
- Generate action tubes.



Main Algorithmic Steps



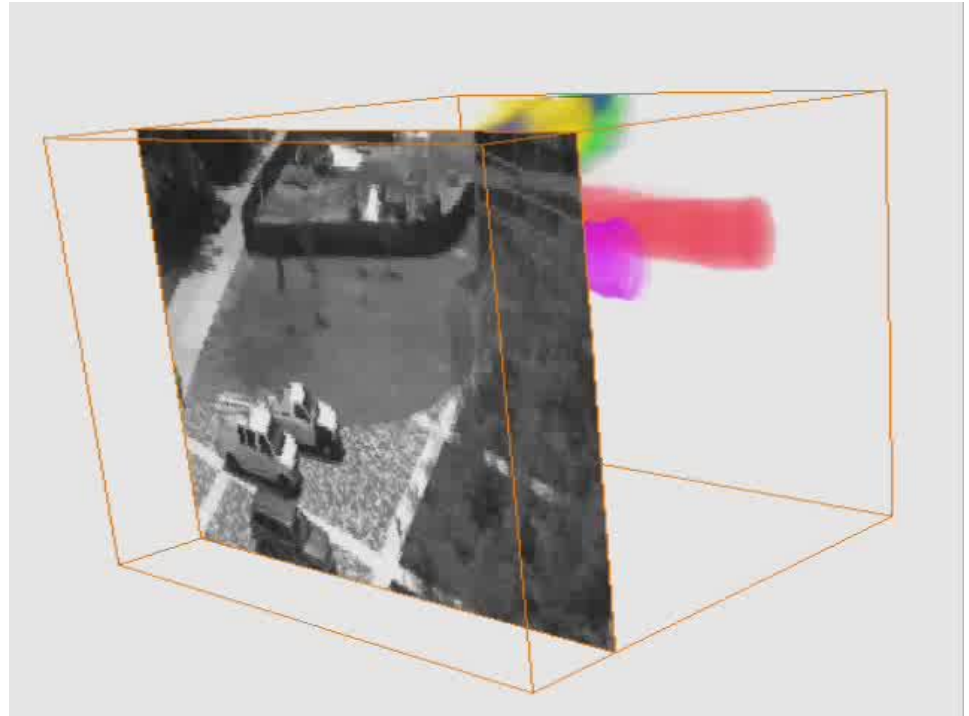
Object Detection

- Background model:
 - Temporal median over 4 minutes of data



Create Activity Tubes

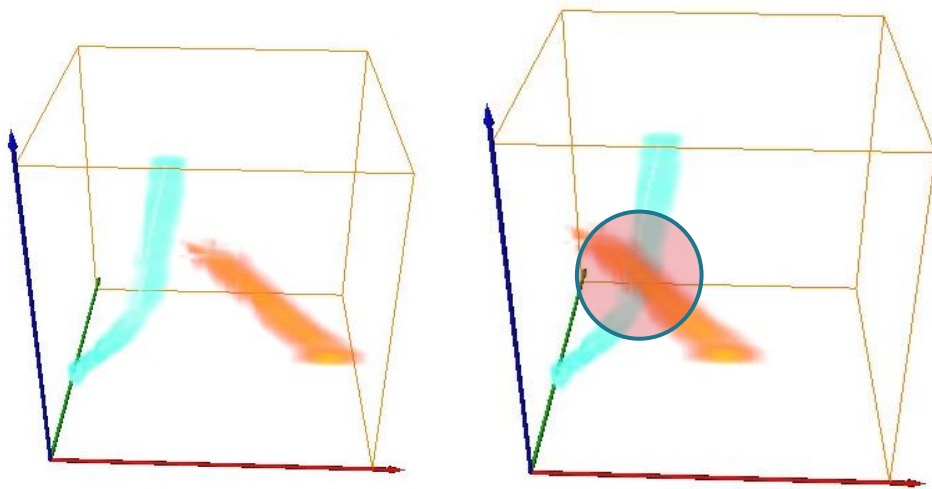
- Trace tracked objects:
 - ▣ Segment tubes
 - ▣ Label tubes



Shift Tubes

□ Energy function:

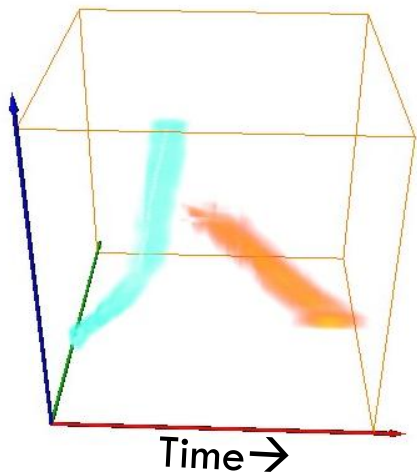
$E = \text{activity cost} + \text{overlap cost} + \text{temporal consistency cost}$



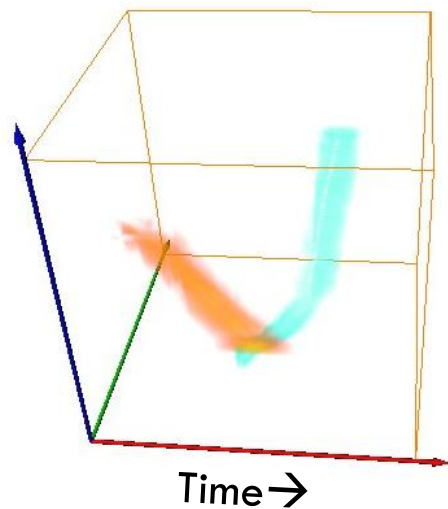
Shift Tubes

□ Energy function:

$E = \text{activity cost} + \text{overlap cost} + \text{temporal consistency cost}$



Blue then orange tube



Orange then blue

A horizontal bar at the top of the slide, divided into a red section on the left and a teal section on the right. The text 'Video Summarization Results' is written in white on the teal section.

Video Summarization Results

Results



Source: www.cs.huji.ac.il/~yaelpri/

Results

SURVEILLANCE SYNOPSIS

Original Video	Surveillance Synopsis (3 second)	Surveillance Synopsis (1 second)																
	<table border="1"><tr><td>Captured frames:</td><td>0</td></tr><tr><td>Dropped frames:</td><td>0</td></tr><tr><td>File size:</td><td>0</td></tr><tr><td>Video length:</td><td>0 Seconds</td></tr><tr><td>Capture length:</td><td>0 Seconds</td></tr><tr><td colspan="2">Capture Properties</td></tr><tr><td>Frame size (pixels):</td><td>708 x 260</td></tr><tr><td>Frame rate:</td><td>15.0 frames/sec</td></tr></table>	Captured frames:	0	Dropped frames:	0	File size:	0	Video length:	0 Seconds	Capture length:	0 Seconds	Capture Properties		Frame size (pixels):	708 x 260	Frame rate:	15.0 frames/sec	
Captured frames:	0																	
Dropped frames:	0																	
File size:	0																	
Video length:	0 Seconds																	
Capture length:	0 Seconds																	
Capture Properties																		
Frame size (pixels):	708 x 260																	
Frame rate:	15.0 frames/sec																	

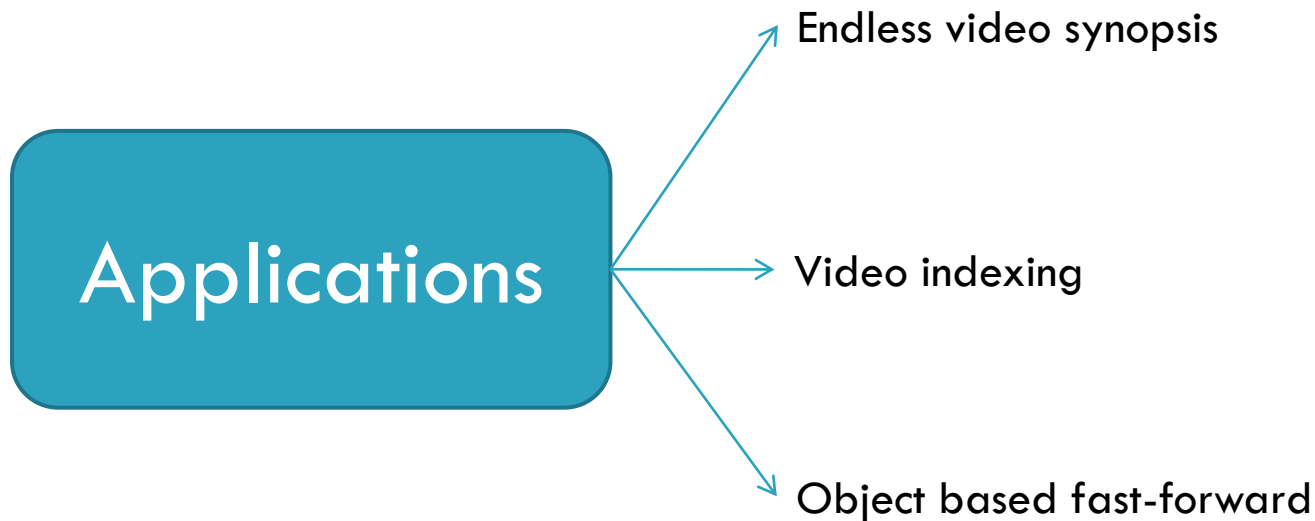
Results



Source: www.cs.huji.ac.il/~yaelpri/

Other Applications

- Using the same principles:



Can we do better?



- Can we avoid background subtraction?
- Can we avoid tracking?
- Improve computational efficiency
- New application domains:
 - ▣ UAV Videos, P-T-Z Surveillance Cameras etc

Thank you



Source: www.cs.huji.ac.il/~yaelpri/