NON-CHRONOLOGICAL VIDEO SYNOPSIS AND INDEXING

Yael Pritch Alex Rav-Acha Shmuel Peleg

Presented by *Mikel Rodriguez*

Problem Introduction

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

Objective

- Generate a video summary

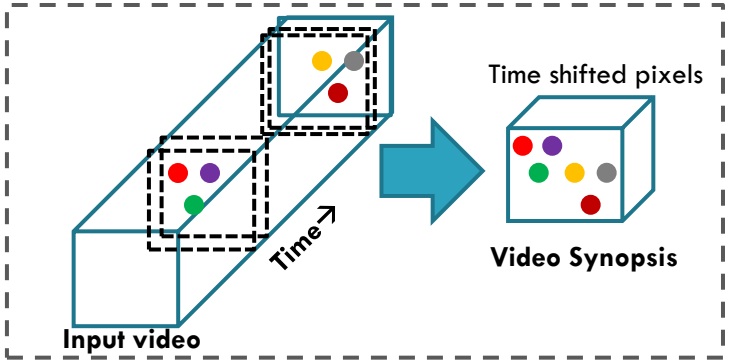
Input: Long video

Output: Short video

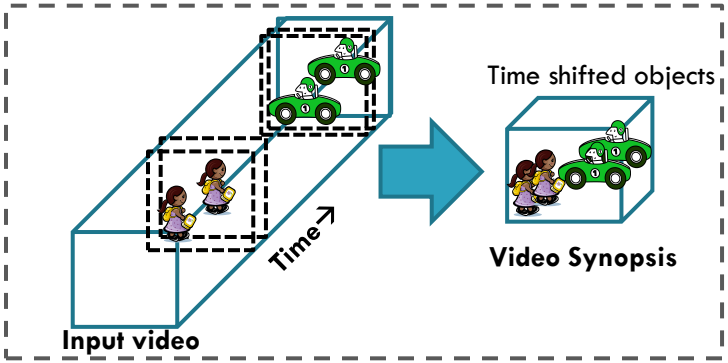
Source: www.cs.huji.ac.il/~yaelpri/

Approaches Overview

Low-level video summary



Object-based video summary



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:

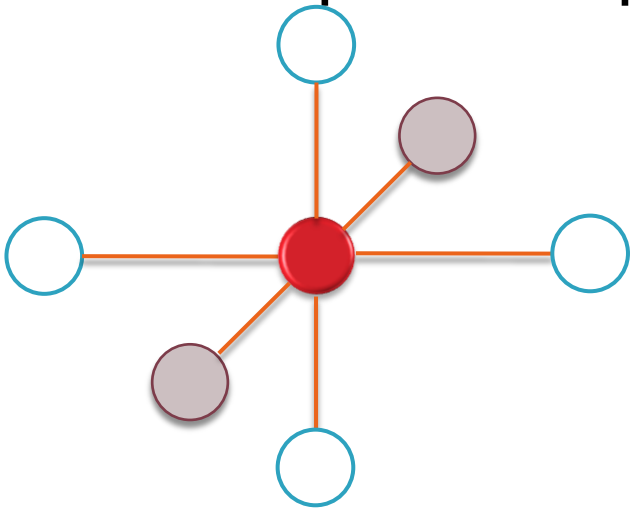
$$\chi(x, y, t) = \| I(x, y, t) - B(x, y, t) \|$$

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:

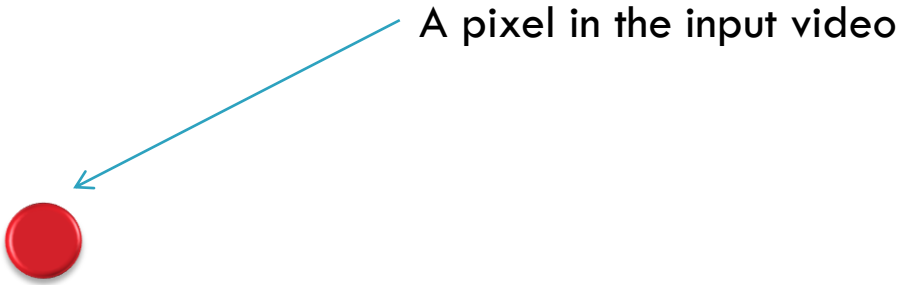


4 spatial neighbors

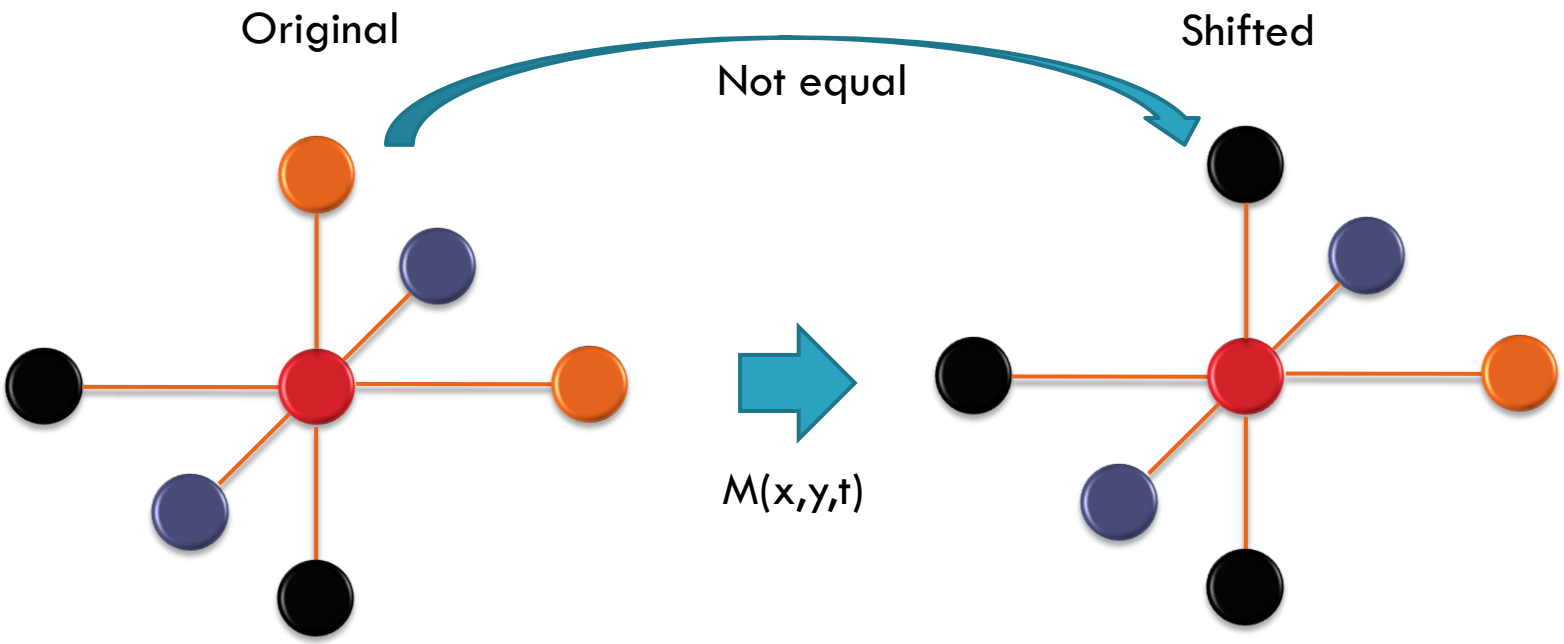
2 temporal neighbors

Discontinuity Cost

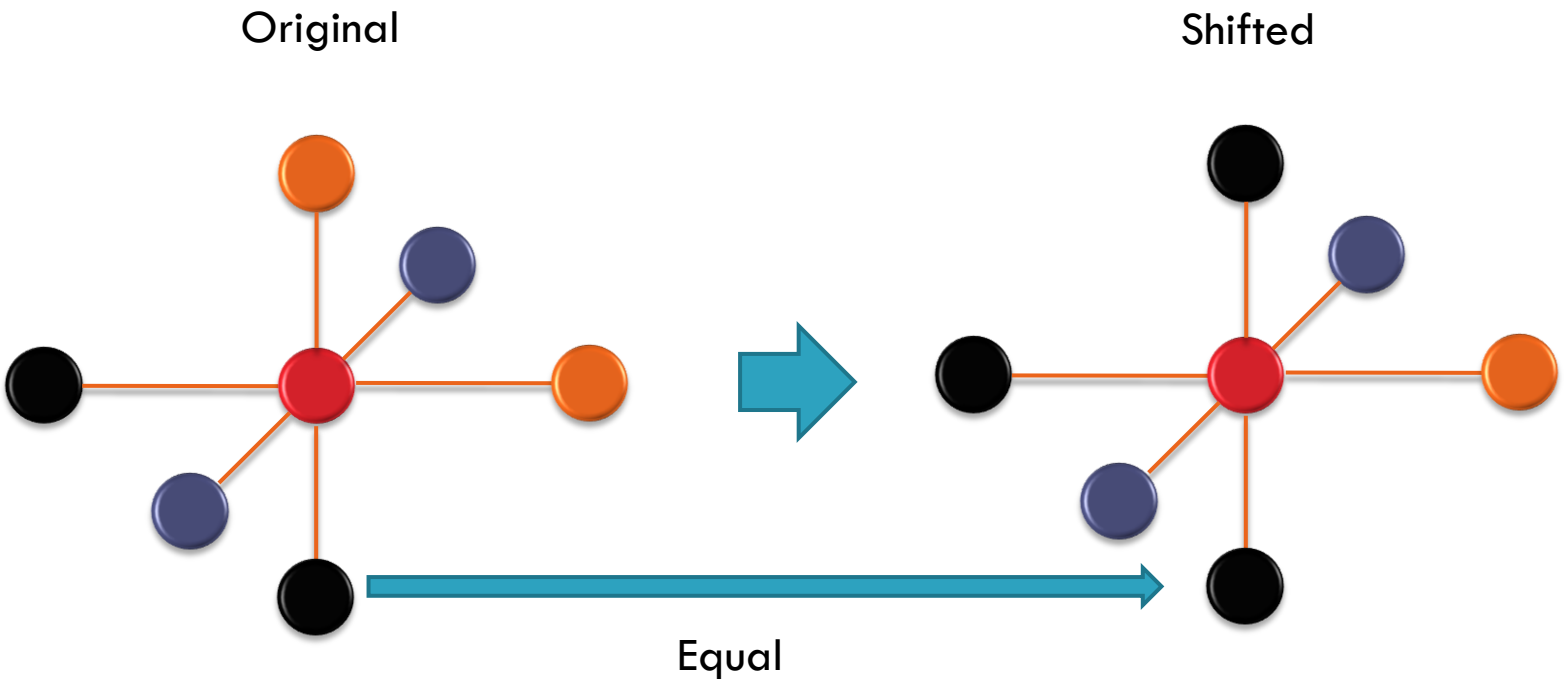
Original



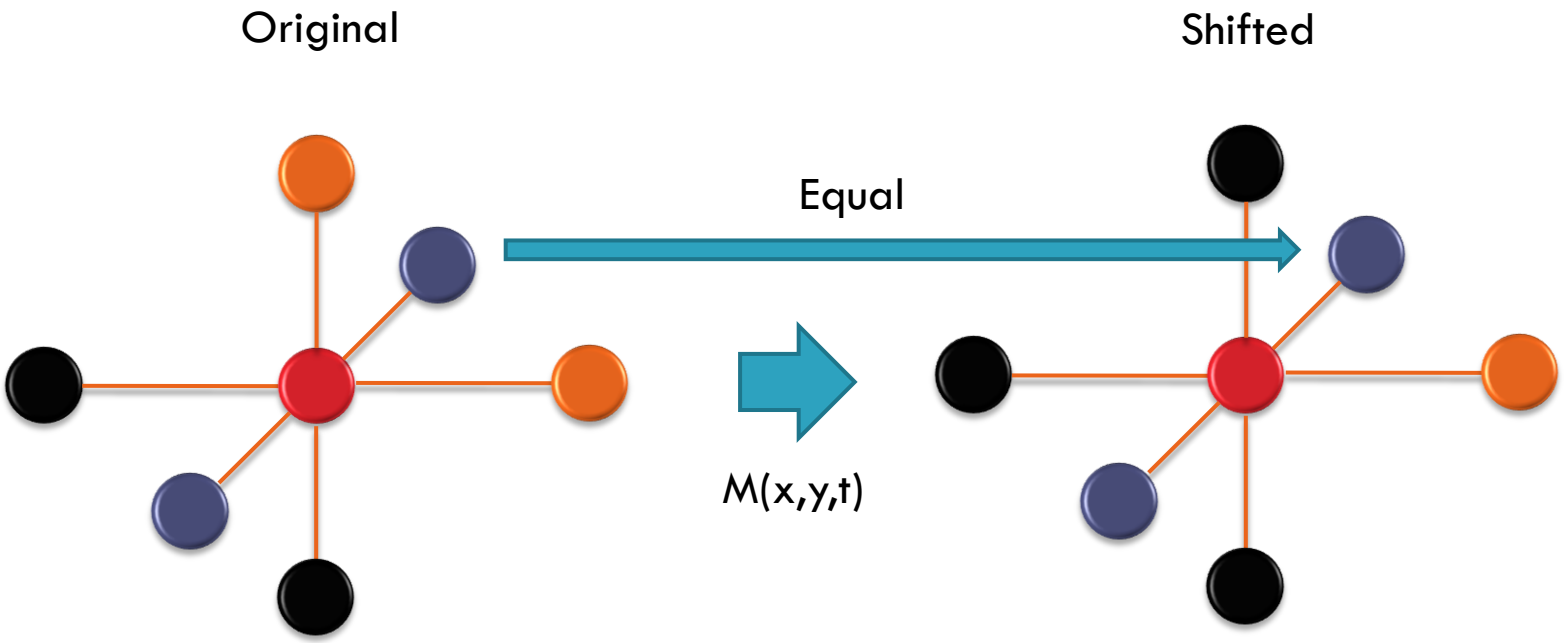
Discontinuity Cost



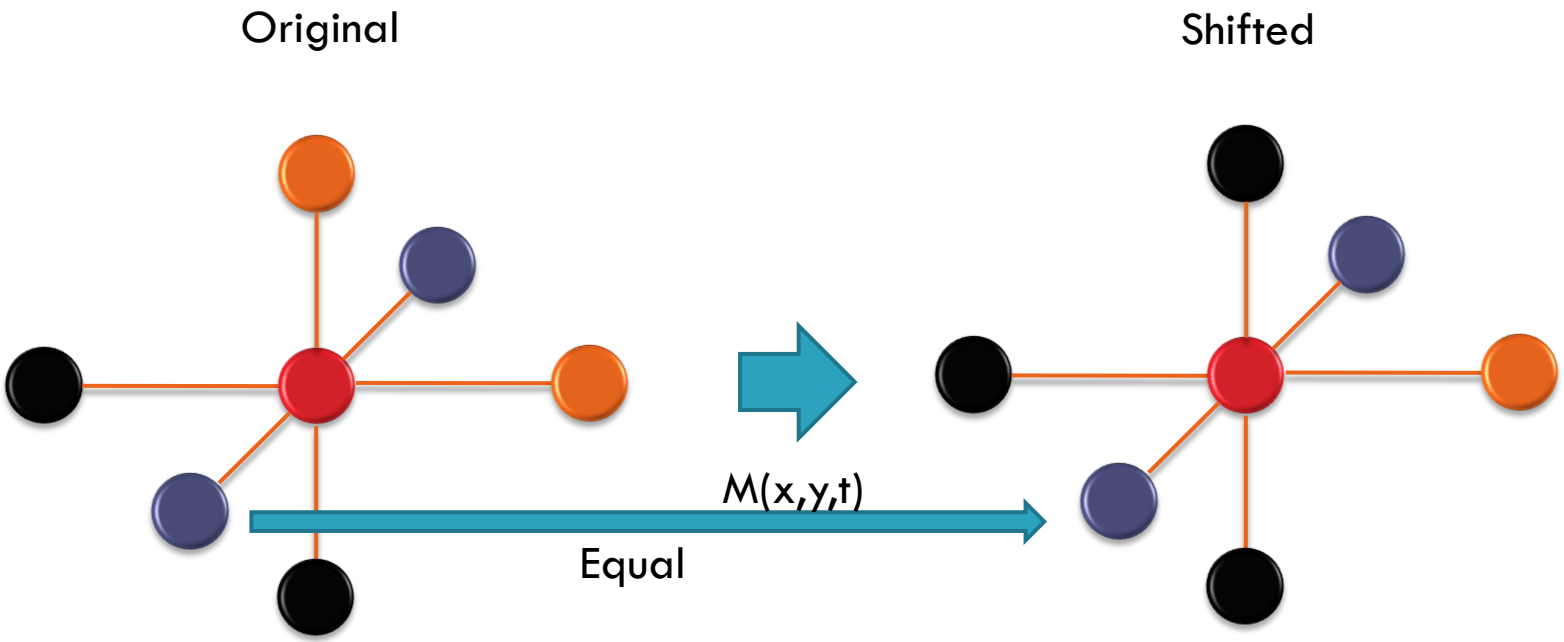
Discontinuity Cost



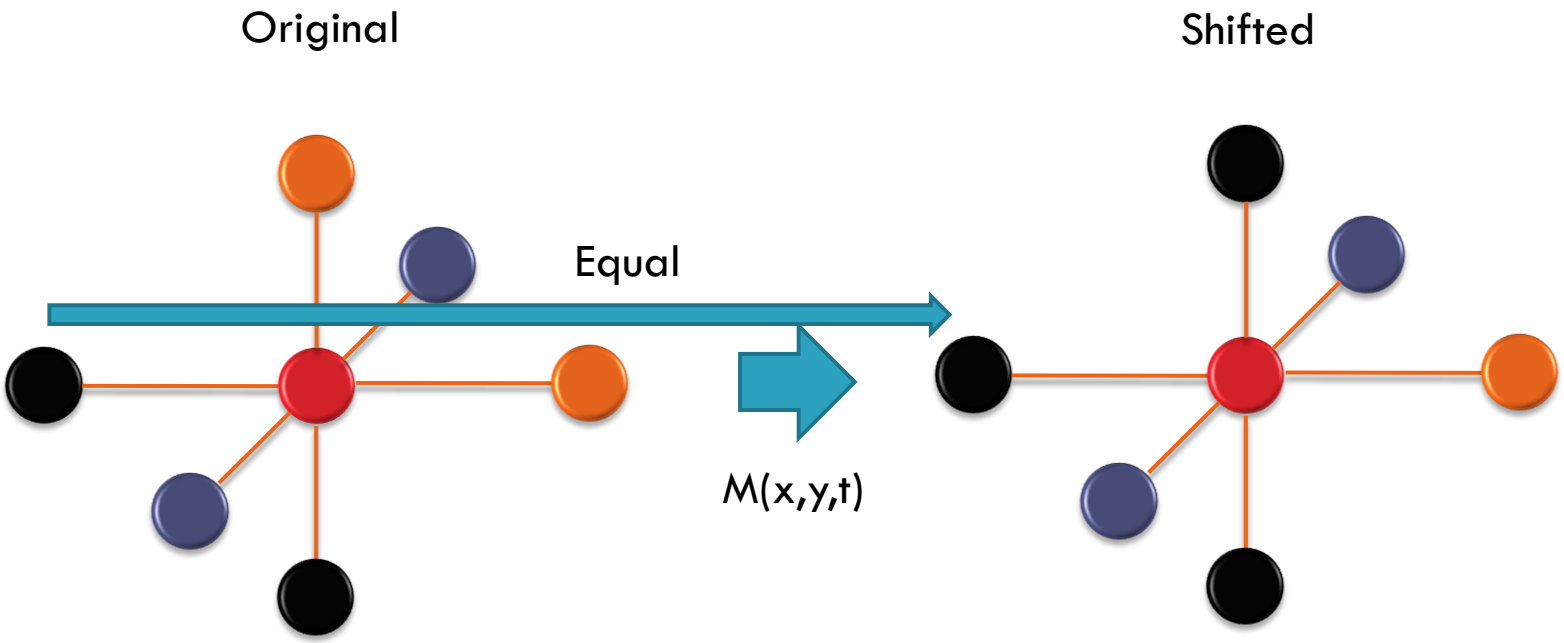
Discontinuity Cost



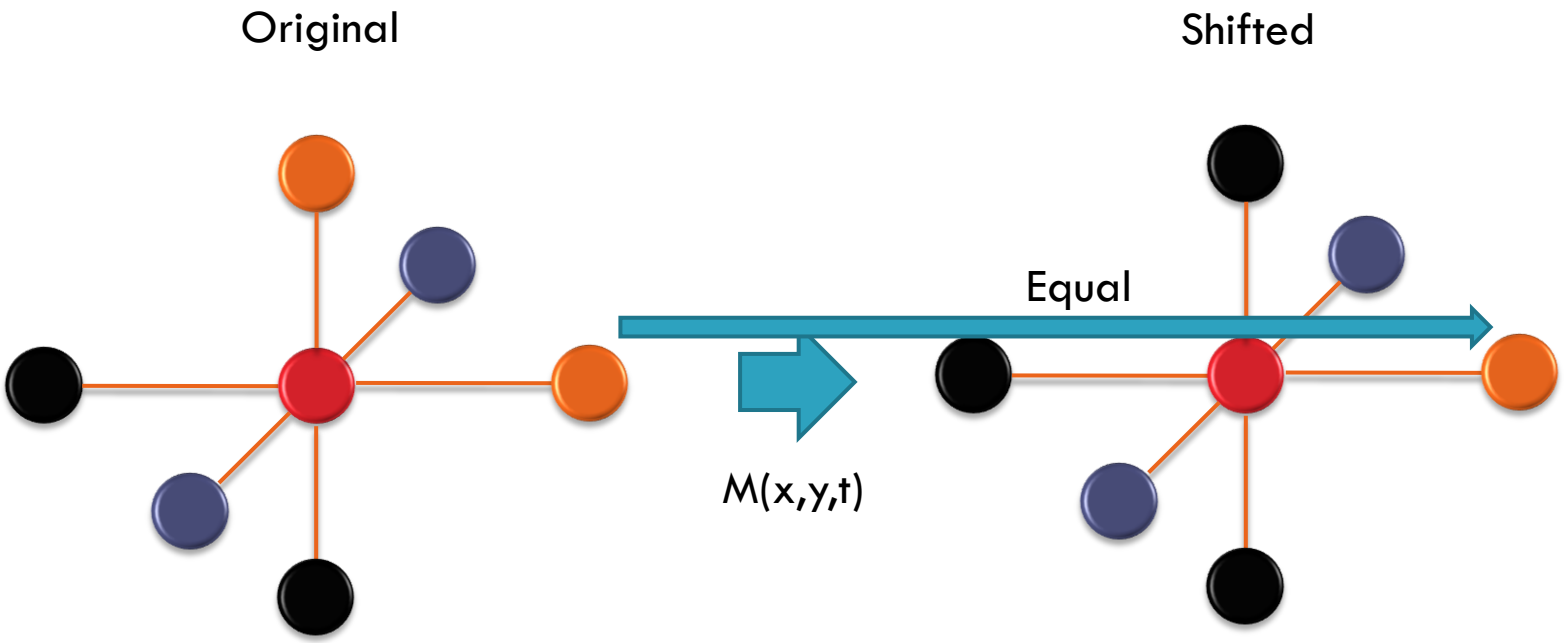
Discontinuity Cost



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

Object-based Video Synopsis

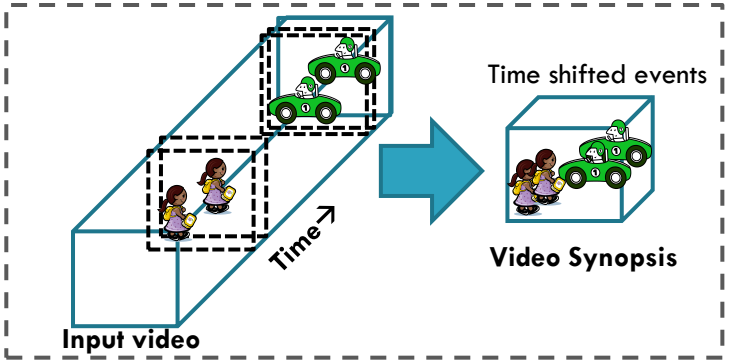
Object-based Video Synopsis

Objective:

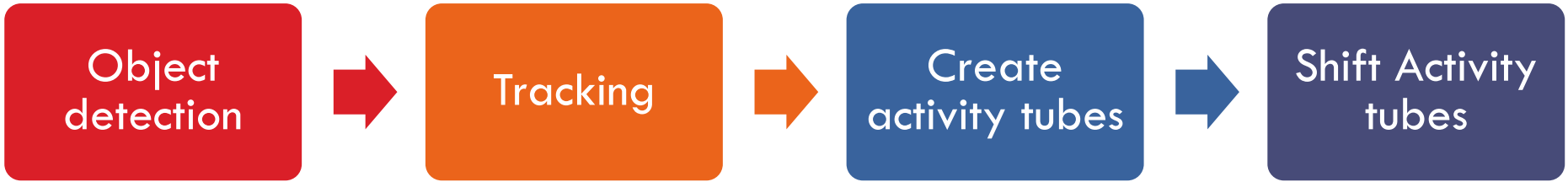
- Create synopsis of a video
  - Combine activities from different times

Approach:

- Object detection and tracking.
- Generate action tubes.



Main Algorithmic Steps

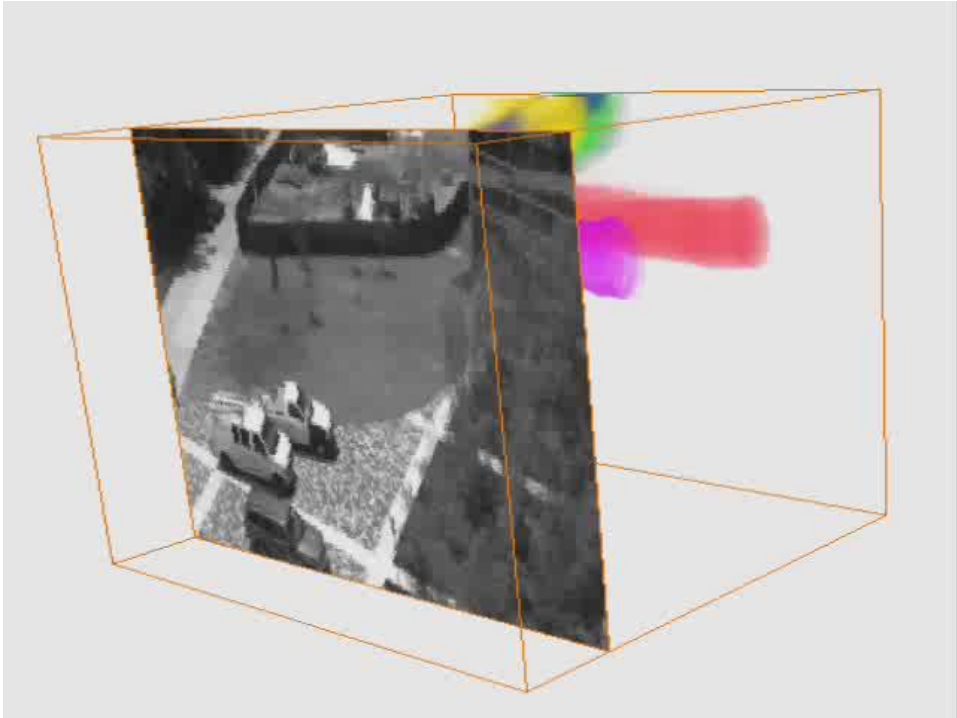


Object Detection

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

Create Activity Tubes

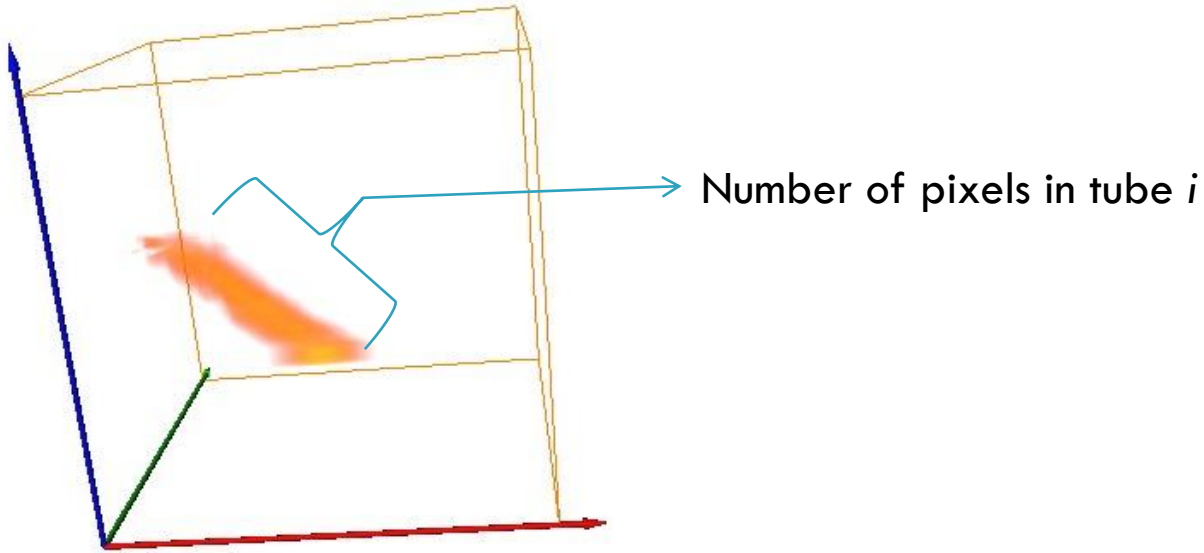
- Trace tracked objects:
  - Segment tubes
  - Label tubes



Shift Tubes

- Energy function:

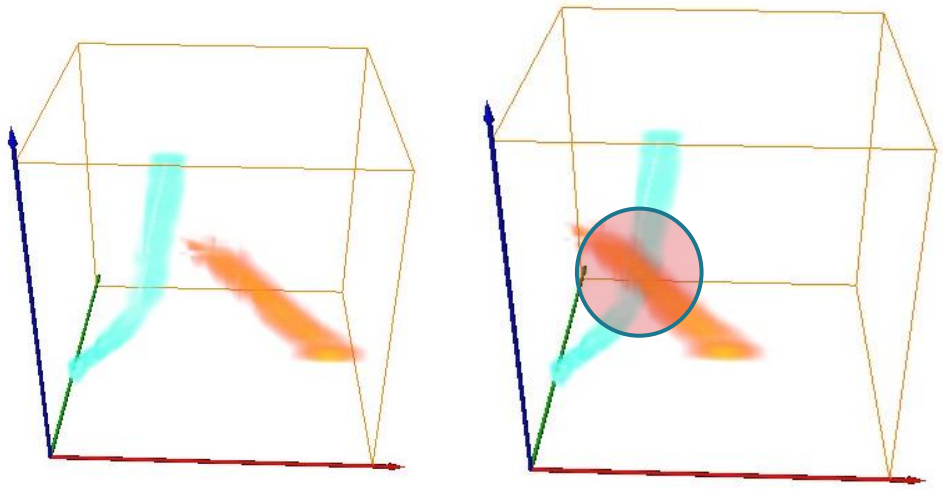
  E = **activity cost** + overlap cost + temporal consistency cost



Shift Tubes

- Energy function:

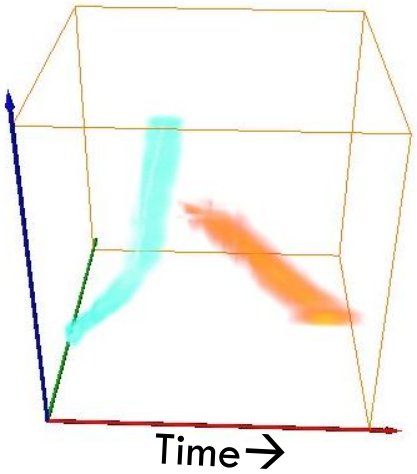
  E = activity cost + **overlap cost** + temporal consistency cost



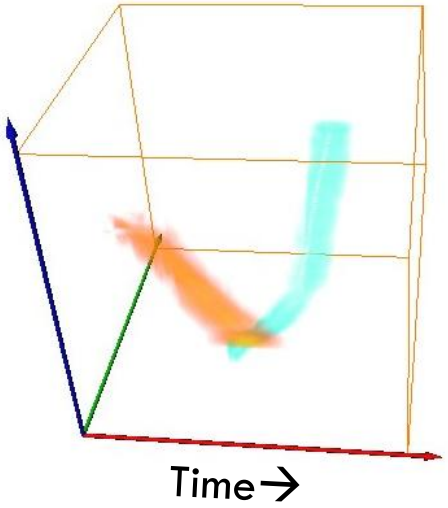
Shift Tubes

- Energy function:

  E = activity cost + overlap cost + **temporal consistency cost**



Blue then orange tube



Orange then blue

Video Summarization Results

Results

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

Results

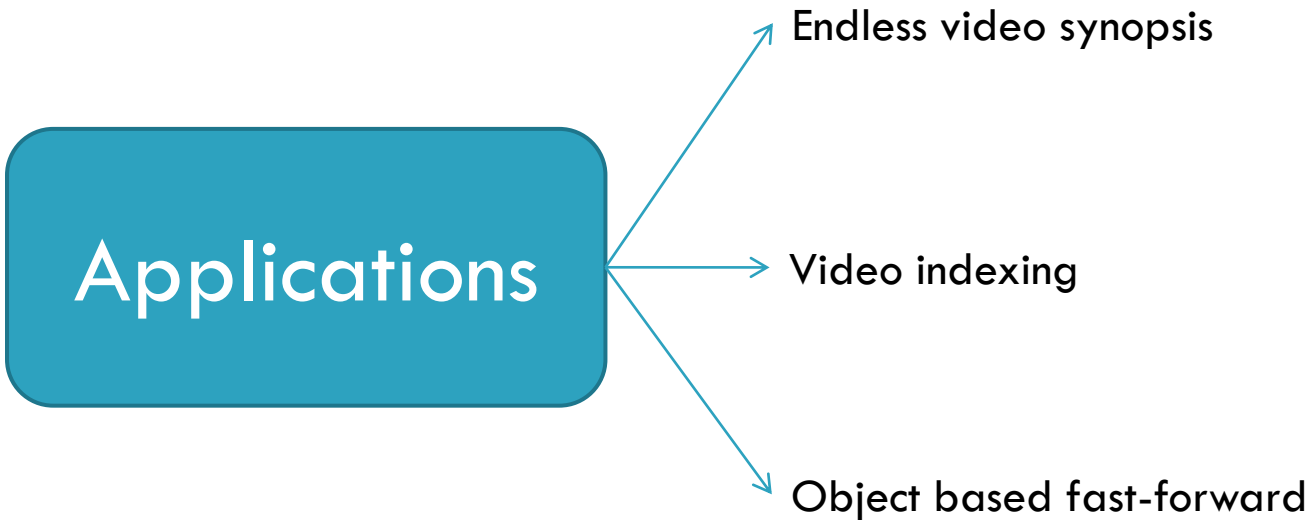
Source: www.cs.huji.ac.il/~yaelpri/

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/