

# SurVizor: Visualizing and Understanding the Key Content of Surveillance Videos

Guodao Sun, Tong Li, Ronghua Liang

Reporter: Tong Li  
Zhejiang University of Technology

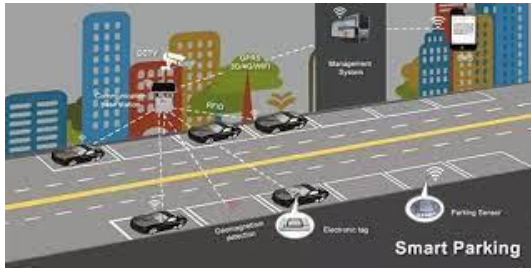
# CONTENT

- **Background**
- **Approaches**
- **Case Study**
- **Discussion**



➤ Application

❑ Intelligent Traffic



❑ Security Monitoring



Surveillance Cameras

● 567 mill. 2021

● 349 mill. 2018

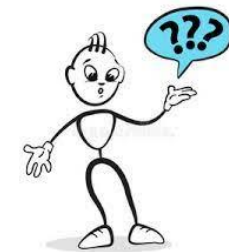
● 176 mill. 2015

➤ Challenge

❑ Retrieval



❑ Complete Picture



❑ Key Content



=



+



**Multi-feature**

+

**Time Series**

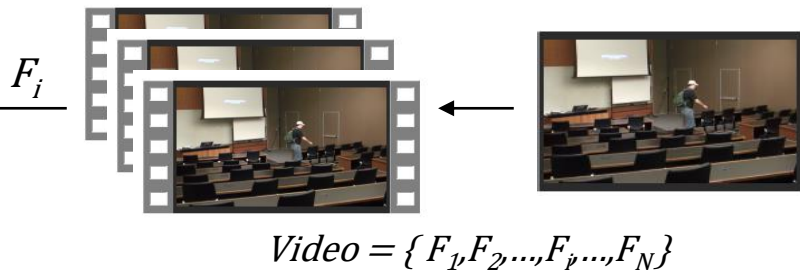
+

**Visual Analysis System**

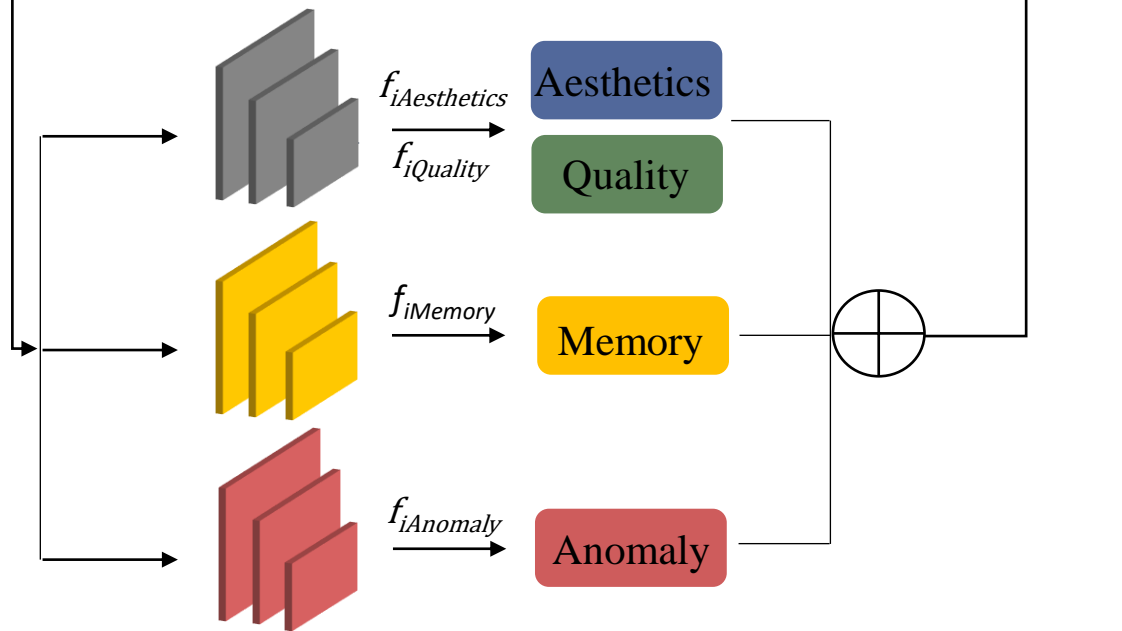
=

**SurVizor**

## Raw Data

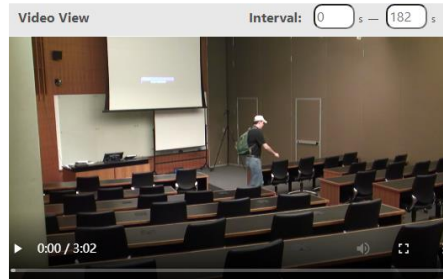


## Data Processing

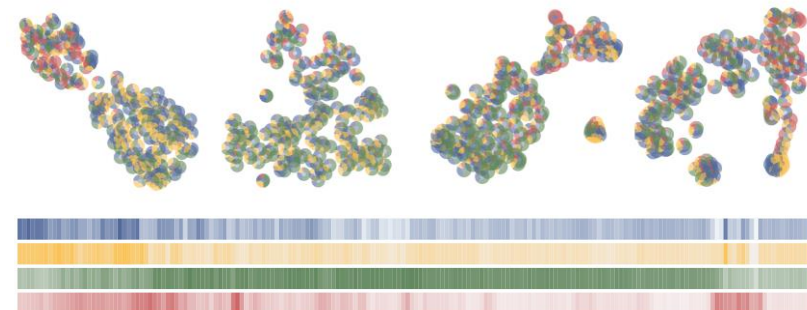


## Video Level

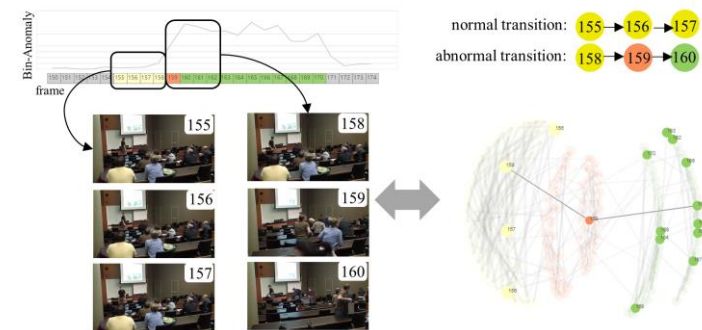
Video	Description	Summary
View HC3	Front view of the lobby	
View HC4	Left view of the lobby	
View IP2	Rear view of the classroom	
View IP5	Front view of the classroom	
Aircraft	An aircraft is landing on the runway	



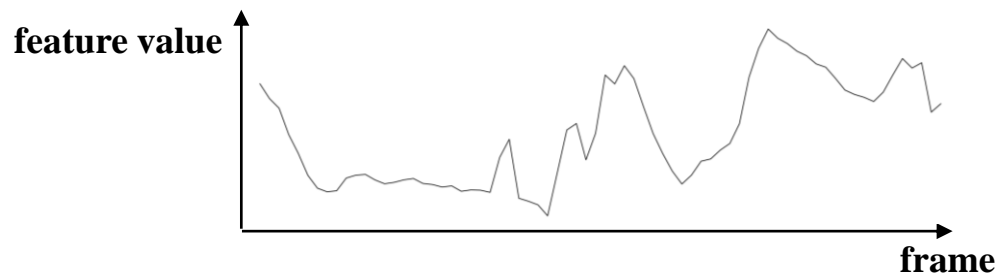
## Feature Level



## Temporal Level



- Quantification ➤ Markov Matrix ➤ Markov Transition Field ➤ Time Series Network



$Q$  bins



$$W = \begin{matrix} & q_1 & q_2 & \dots & q_{O-1} & q_O \\ \begin{matrix} q_1 \\ q_2 \\ \dots \\ q_O \end{matrix} & \begin{pmatrix} 5 & 1 & \dots & 0 & 0 \\ 3 & 4 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 2 & 6 \end{pmatrix} \end{matrix}$$

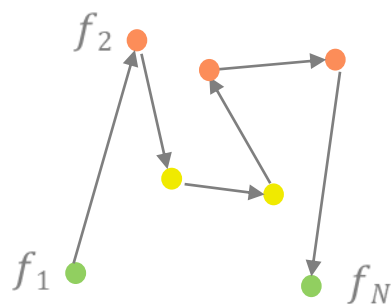
Markov Matrix

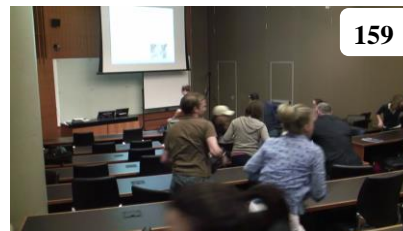
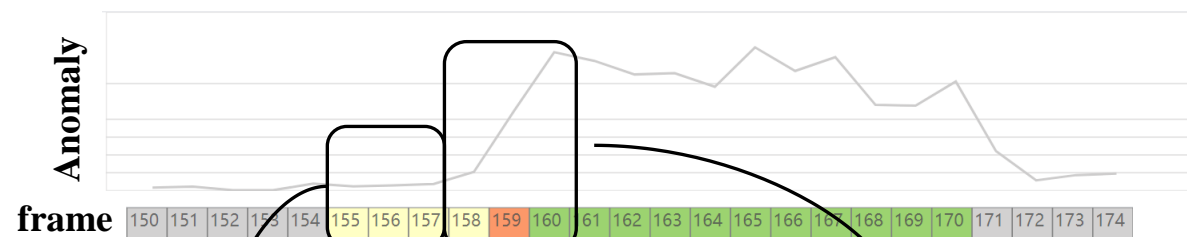
$$W' = \begin{matrix} & q_1 & q_2 & \dots & q_{O-1} & q_O \\ \begin{matrix} q_1 \\ q_2 \\ \dots \\ q_O \end{matrix} & \begin{pmatrix} 0.5 & 0.1 & \dots & 0 & 0 \\ 0.3 & 0.4 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0.2 & 0.6 \end{pmatrix} \end{matrix}$$

MTF

$$M = \begin{matrix} & f_1 & f_2 & \dots & f_{N-1} & f_N \\ \begin{matrix} q_2 \\ f_1 \\ f_2 \\ \dots \\ f_N \end{matrix} & \begin{pmatrix} 0.2 & 0.4 & \dots & 0.3 & 0.4 \\ 0.3 & 0.5 & \dots & 0.3 & 0.5 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0.3 & 0.5 & \dots & 0.3 & 0.5 \end{pmatrix} \end{matrix}$$

Network + Louvain

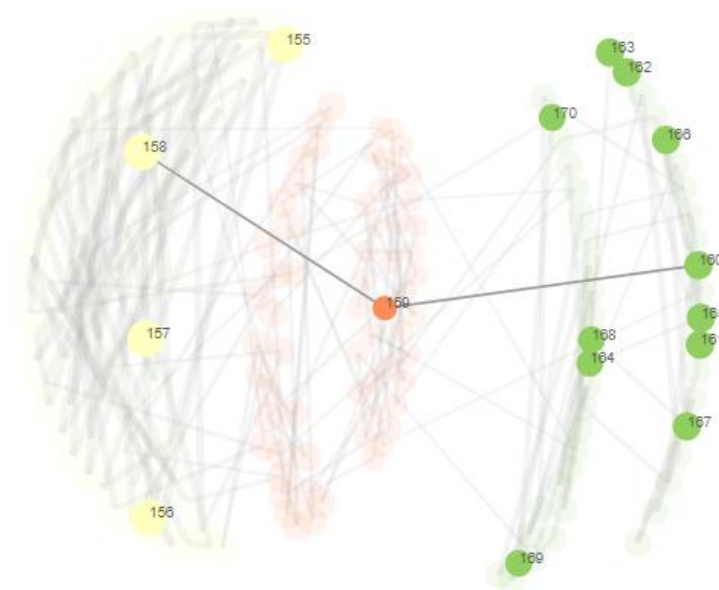




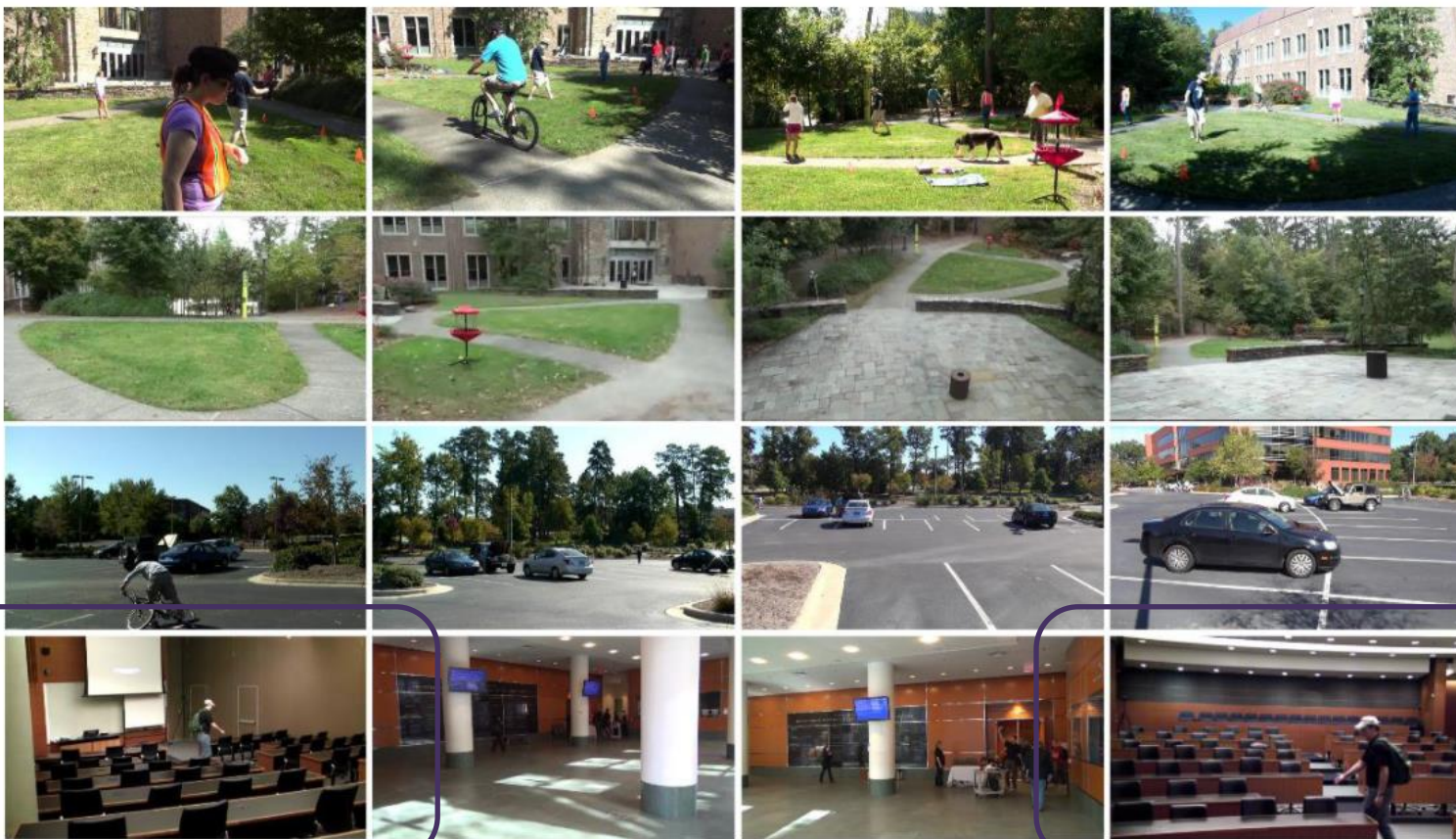
normal transition:



abnormal transition:



➤ **Dataset: Real-world + Multi-camera**



**CAMPUS**

**IP2**

**IP5**

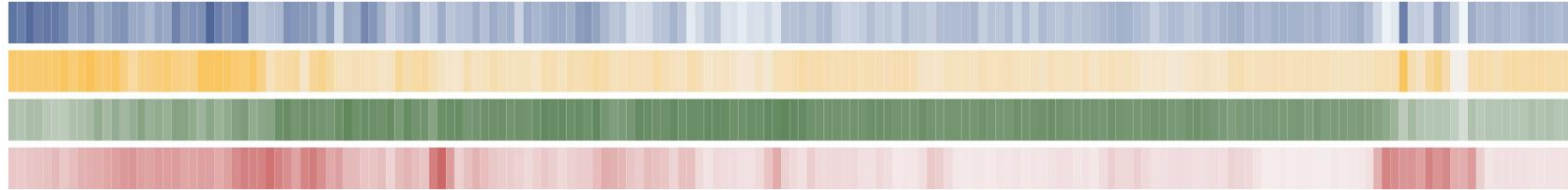
*"Multi-view People Tracking via Hierarchical Trajectory Composition", Yuanlu Xu, Xiaobai Liu, Yang Liu, Song-Chun Zhu, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.*



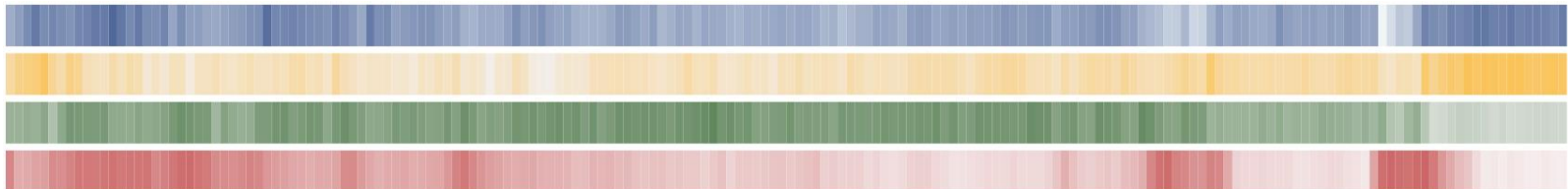
➤ Feature Detail View ( Feature Level )

PS.  Aesthetics  Memory  Quality  Anomaly

IP2

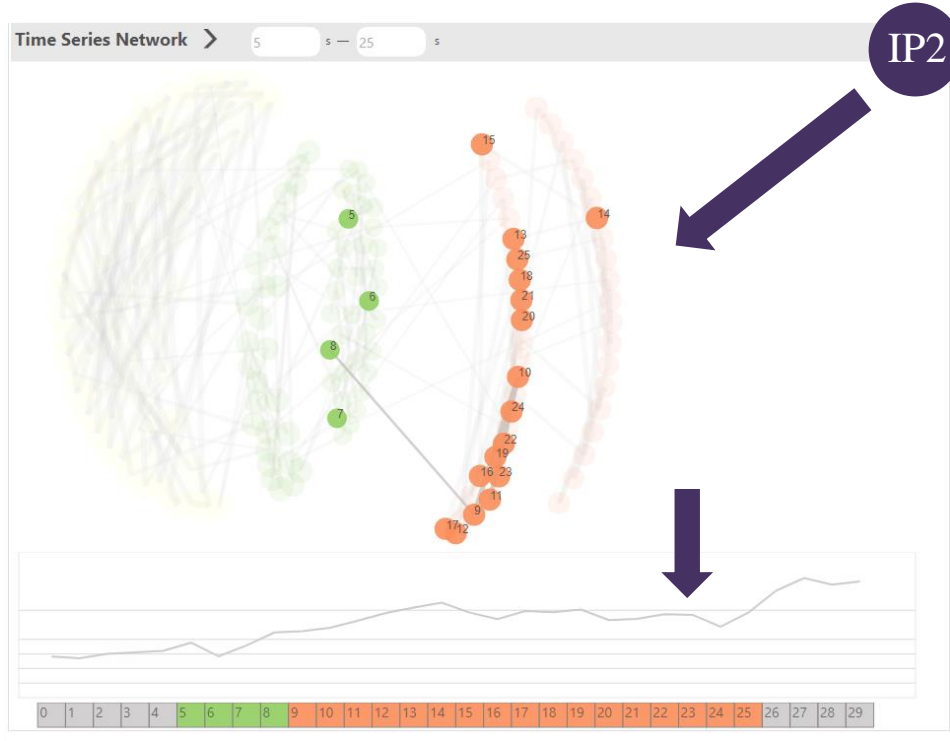


IP5



## ➤ Time Series Network – Anomaly ( Temporal Level )

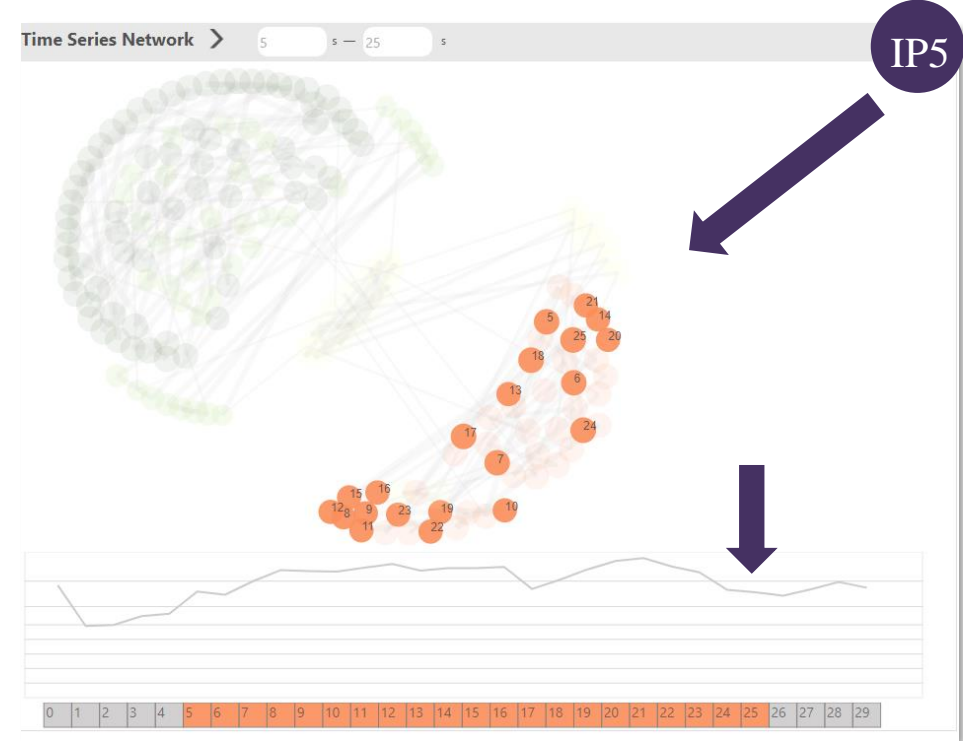
5s – 25s



5



20

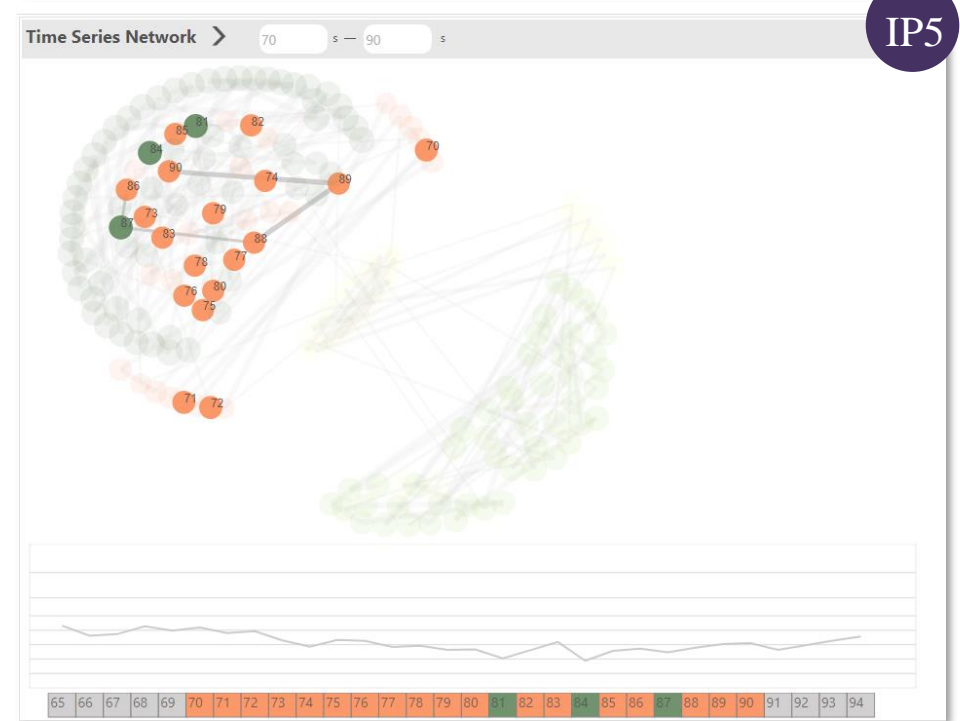


5



20

## ➤ Time Series Network – Anomaly ( Temporal Level ) 70s – 90s



75



85



89



75

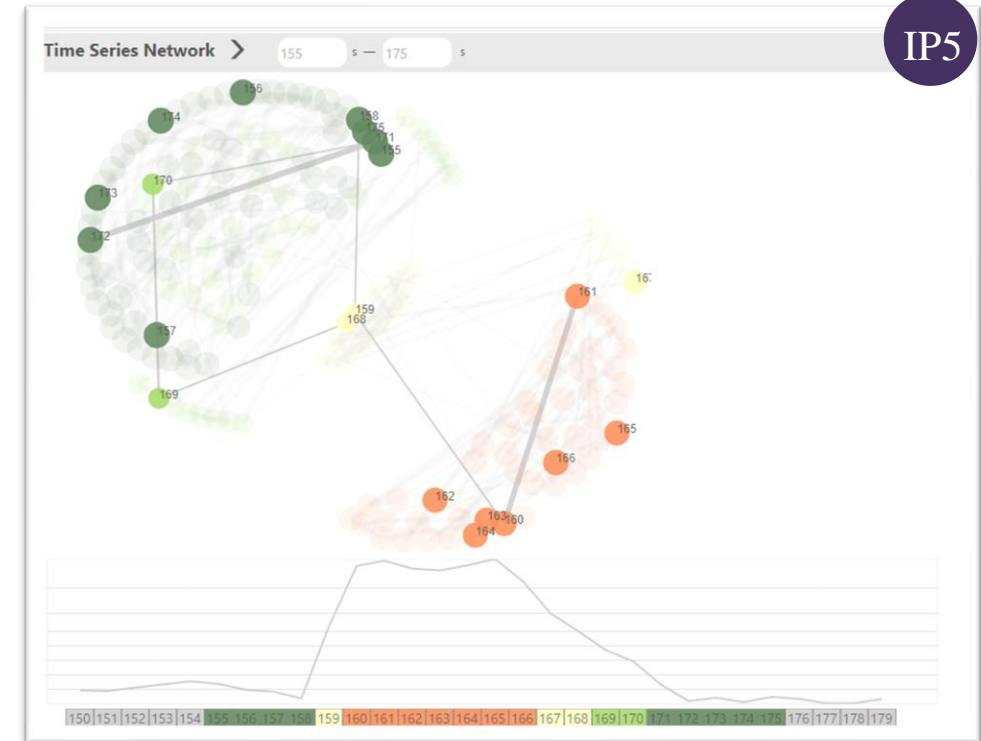
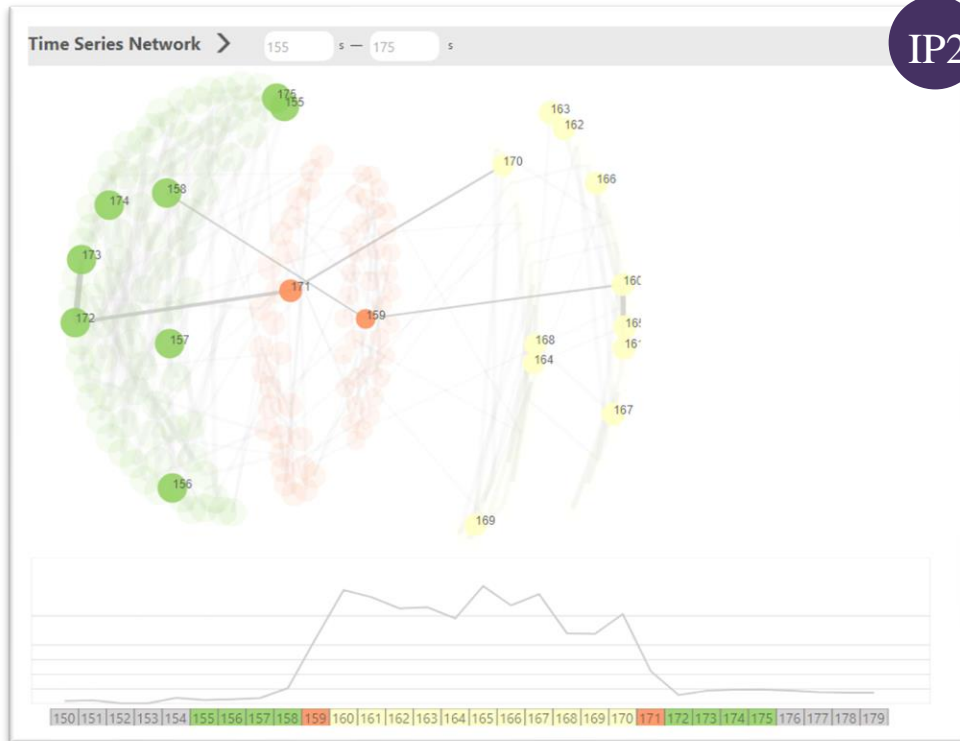


85



89

➤ **Time Series Network – Anomaly ( Temporal Level )**  
**155s – 175s**



155



165



175



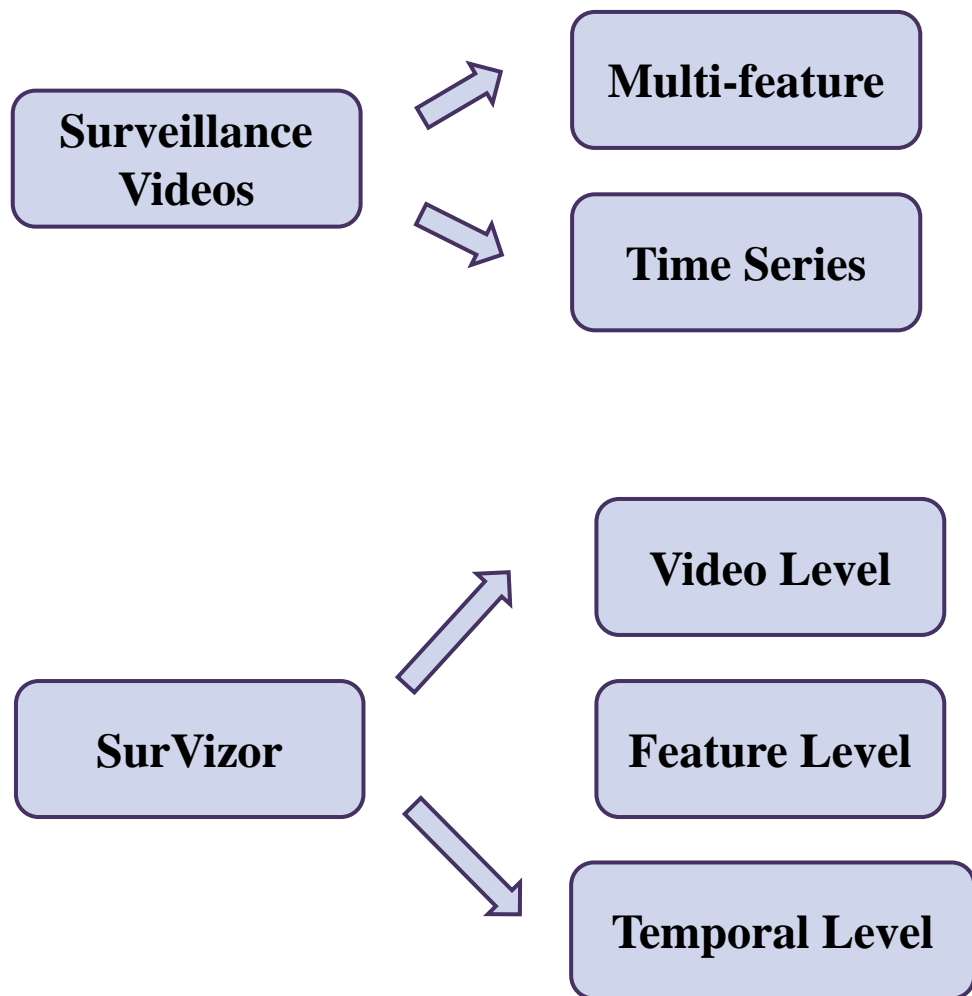
155



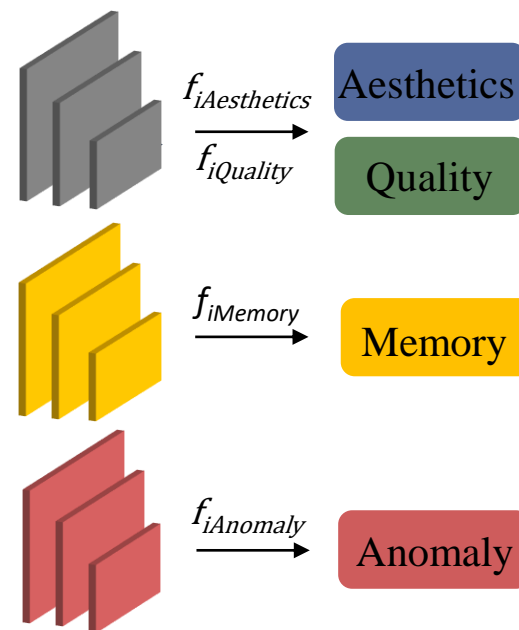
165



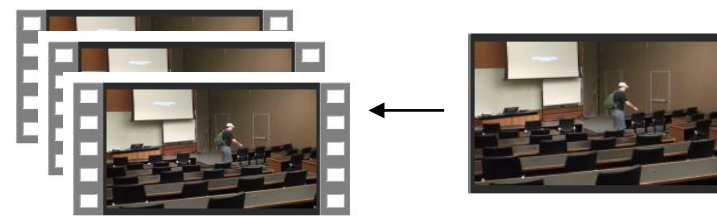
175



### ? Feature Selection



### ? Dynamic Sampling



$$Video = \{F_1, F_2, \dots, F_p, \dots, F_N\}$$

# **SurVizor: Visualizing and Understanding the Key Content of Surveillance Videos**

**Guodao Sun, Tong Li, Ronghua Liang**

---

**Thanks for your attention!**