

Vision – Visualization Communication Report



Tong Li

[TongLi97.github.io](https://github.com/TongLi97)



ZJUTVIS Lab

Zhejiang University of Technology

Outline

- **Video × Vis**

- Pros. and Cons.

- **Related Work**

- OD Vis
- ER Vis
- HPE Vis

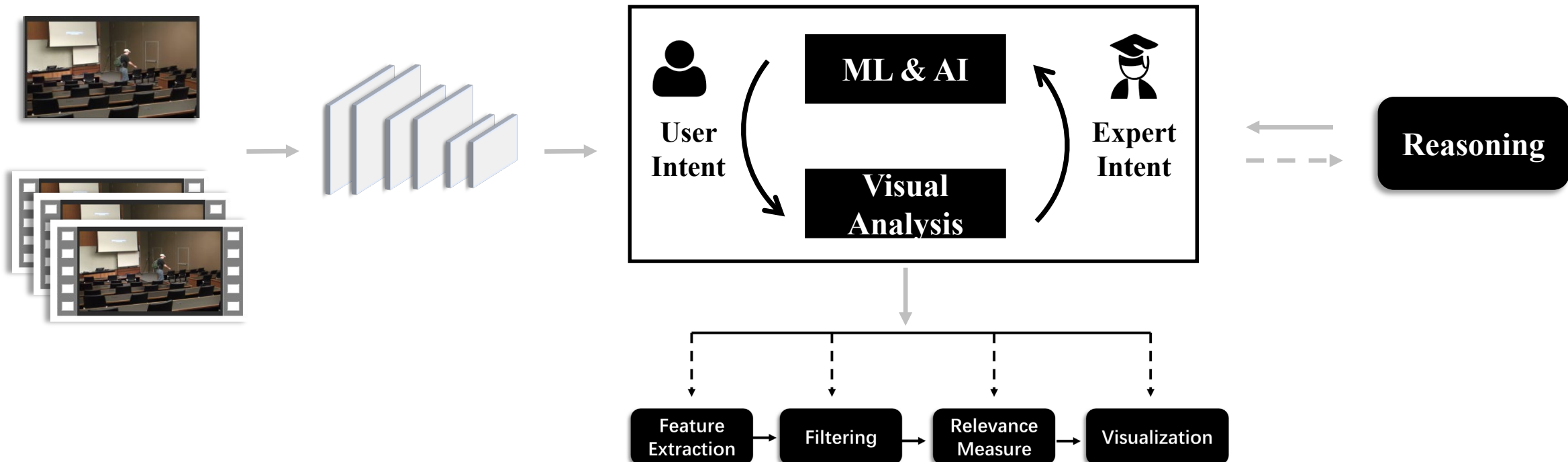
- **Our Work**

- Background and Challenge
- Hierarchical Task Analysis
- Workflow
- Case Demonstration
- Discussion

Pros. and Cons.

- ✓ **Performance:** **Strong** Computational and Comprehension Skills
- × **Manual Inspection:** **Labor-intensive** Tasks
- × **Machine Intelligence:** **Inaccurate** Results

- ✓ **Visibility:** **Improve** the visibility of video content.
- ✓ **Interpretability:** **Improve** the interpretability of video semantic understanding models.



Outline

- Video ✕ Vis

- Pros. and Cons.

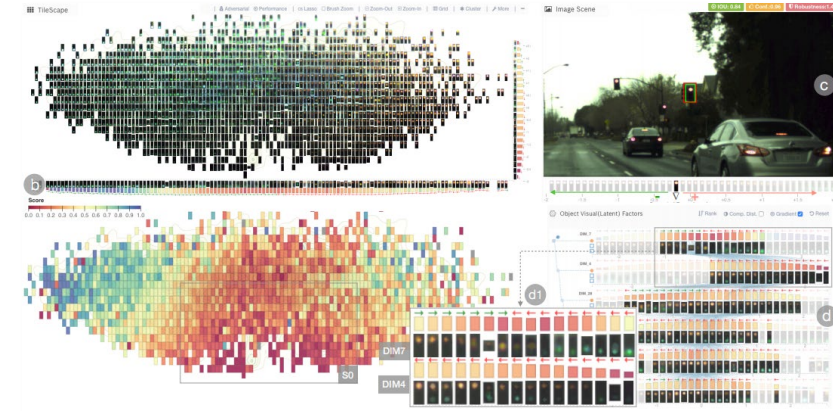
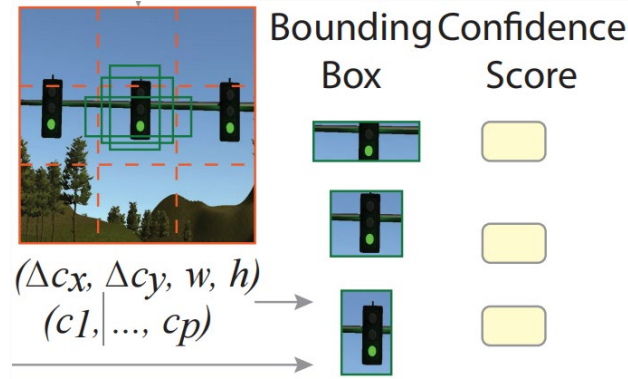
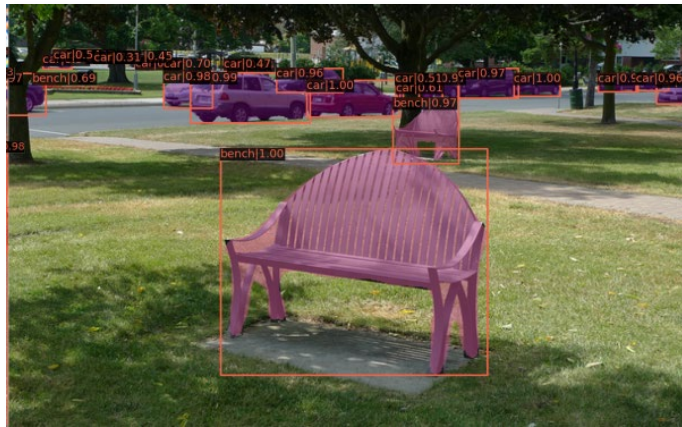
- **Related Work**

- OD Vis
- ER Vis
- HPE Vis

- **Our Work**

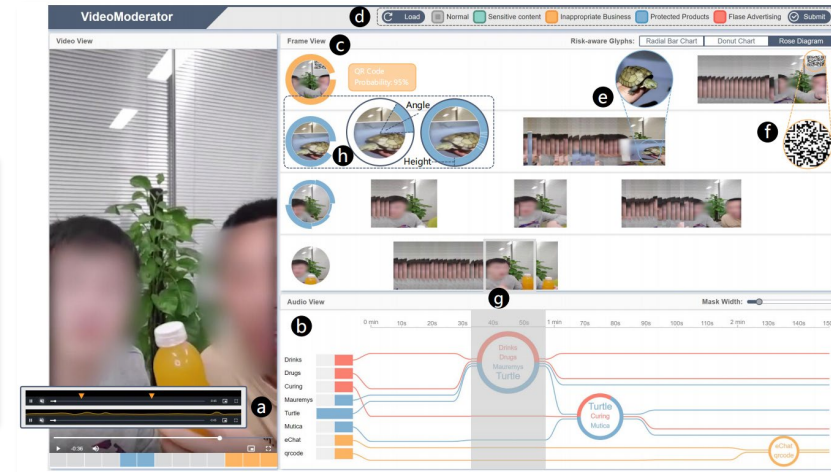
- Background and Challenge
- Hierarchical Task Analysis
- Workflow
- Case Demonstration
- Discussion

Object Detection / Semantic Segmentation ✕ VIS



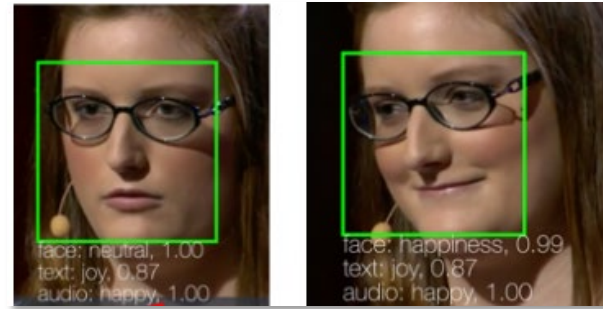
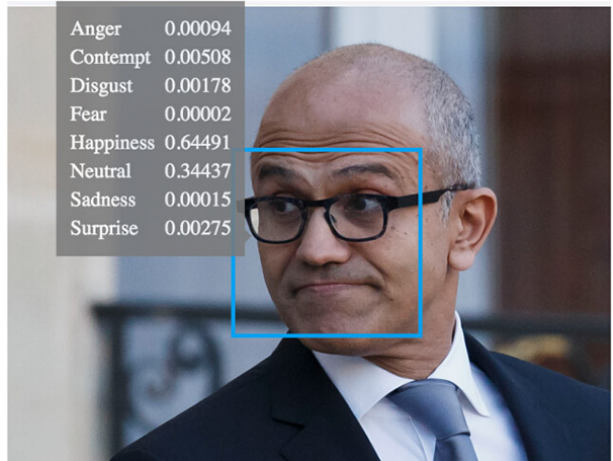
TVCG, 2020 [1]

- ✓ Scene
- ✓ Certain Objects



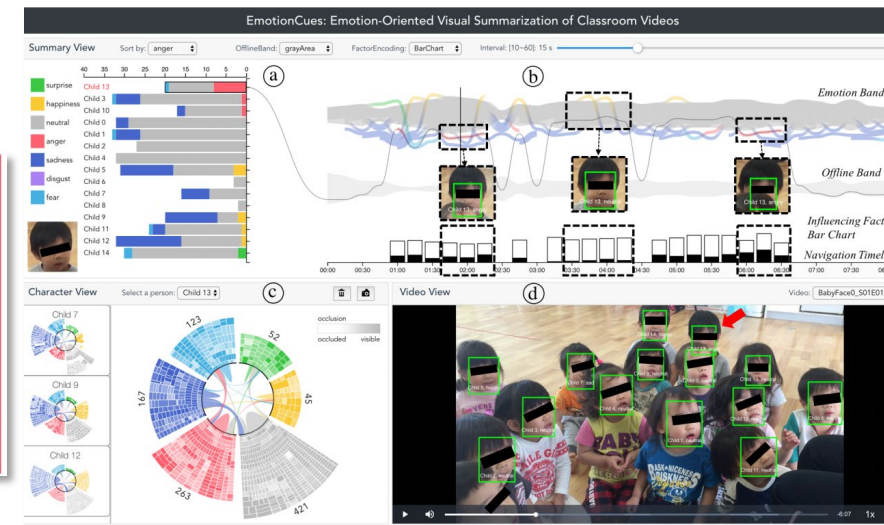
TVCG, 2021 [2]

Emotion Recognition \times VIS



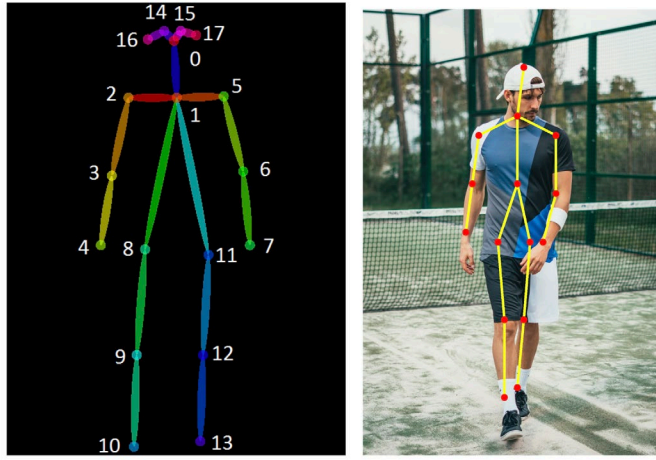
TVCG, 2019 [3]

- ✓ Human Face
- ✓ Clear Shot

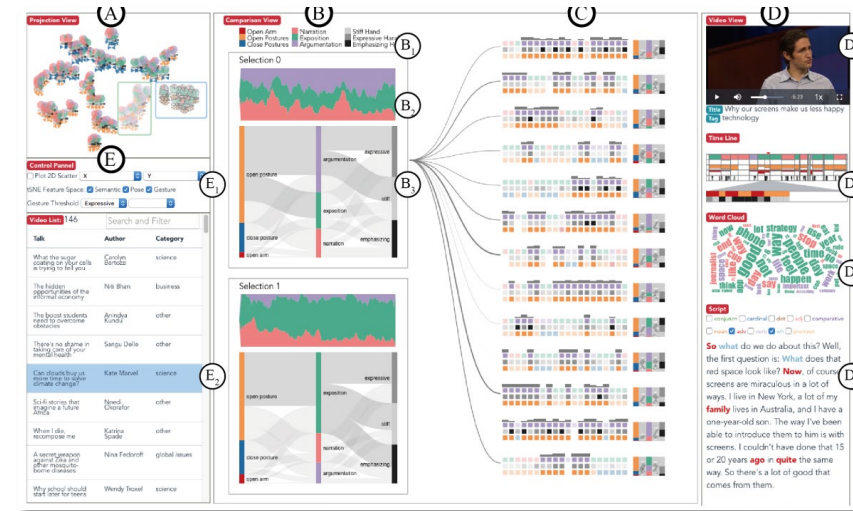
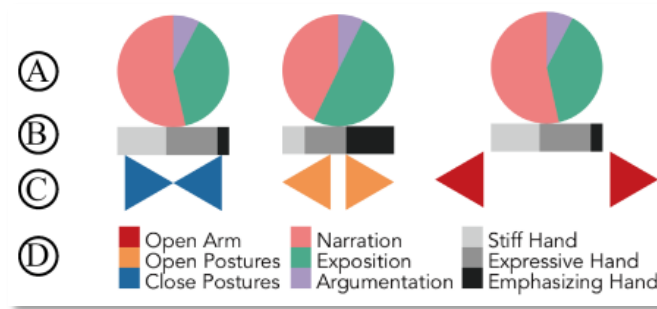


TVCG, 2020 [4]

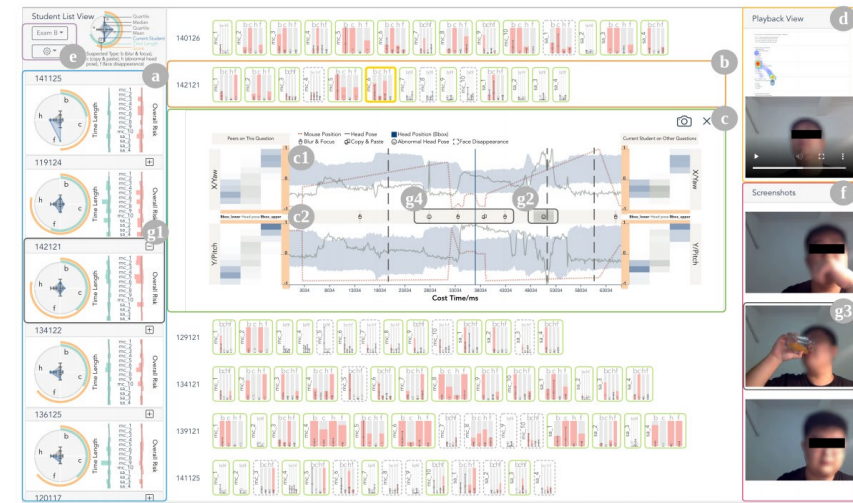
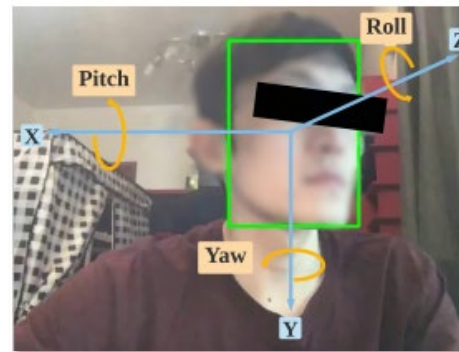
Human Pose Estimation ✕ VIS



- ✓ Skeleton, Keypoints, Pose
- ✓ Body Expression



TCVG, 2018 [5]



CHI, 2021 [6]

Outline

- Video ✕ Vis
 - Pros. and Cons.

- Related Work
 - OD Vis
 - ER Vis
 - HPE Vis

- **Our Work**
 - Background and Challenge
 - Hierarchical Task Analysis
 - Workflow
 - Demo
 - Case Demonstration
 - Discussion

Surveillance Video

□ Data Challenge

Big Data、 Uneven Quality

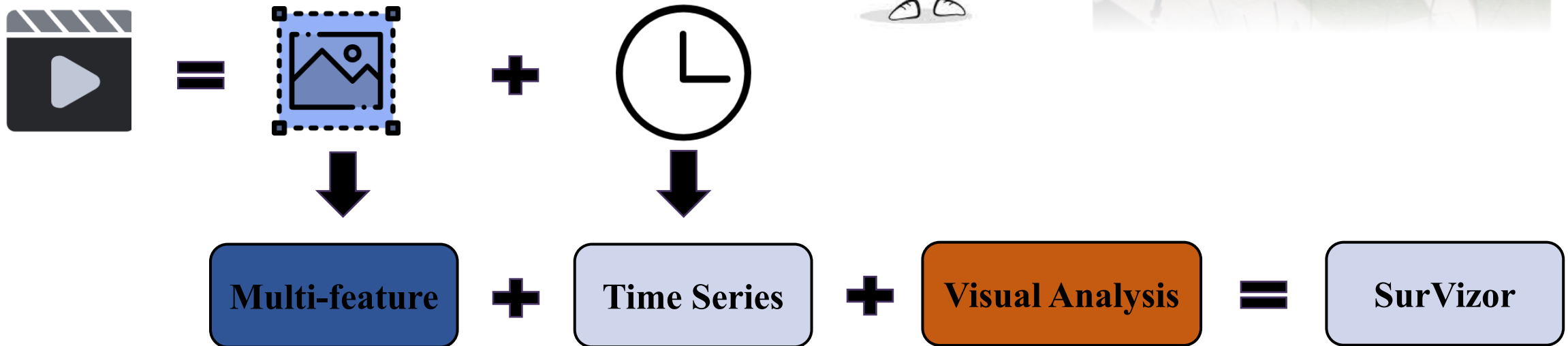
Noise Data

Loose Structures or Without Story Units

□ Analytics Goal

Reduce the time of reviewing videos.

Understand video with low cost.



SurVizor: Hierarchical Task Analysis

□ T1. Data Processing

T1.1. Collect Data

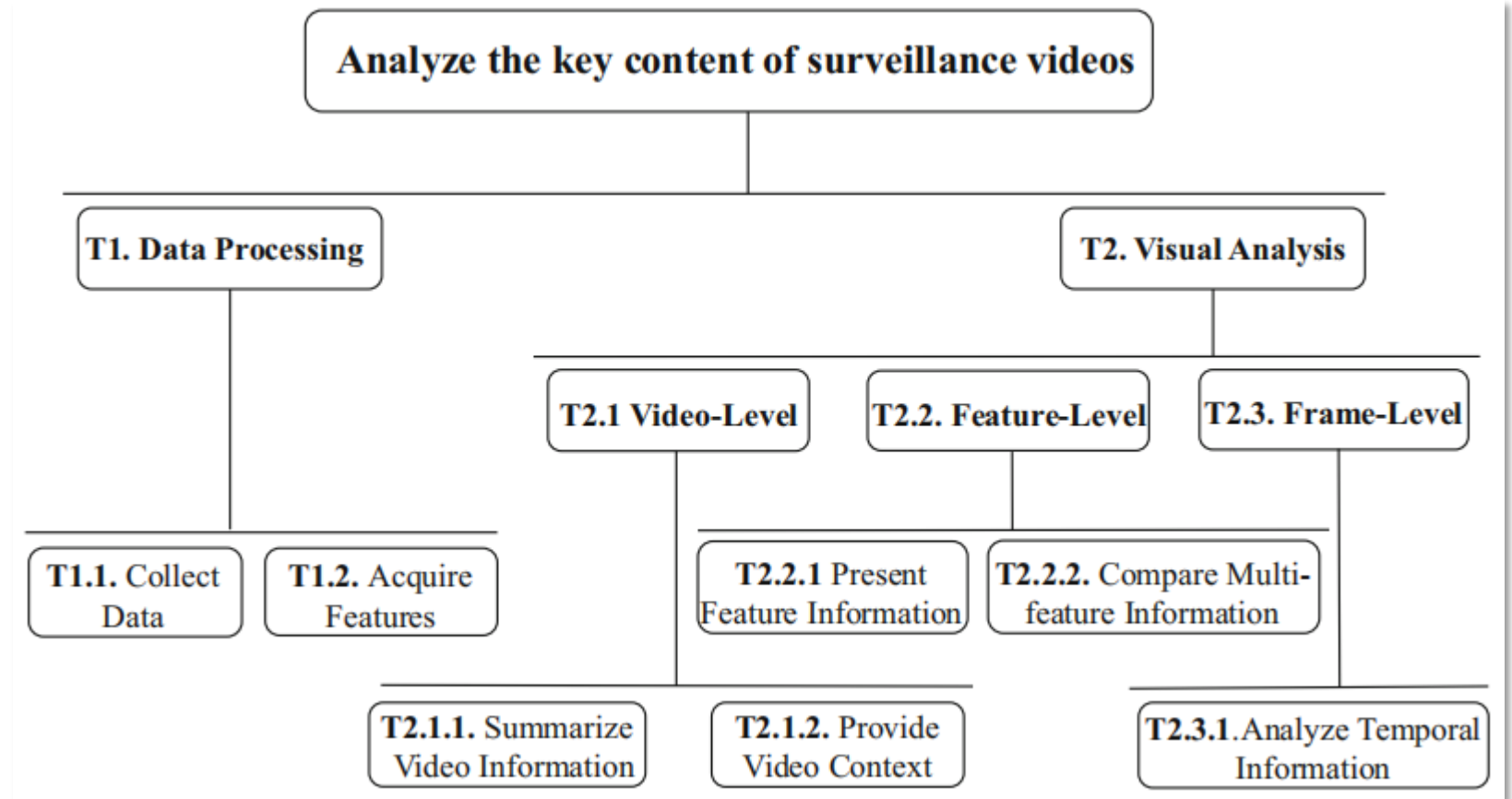
T1.2. Acquire Features

□ T2. Visual Analysis

T2.1. Video-Level

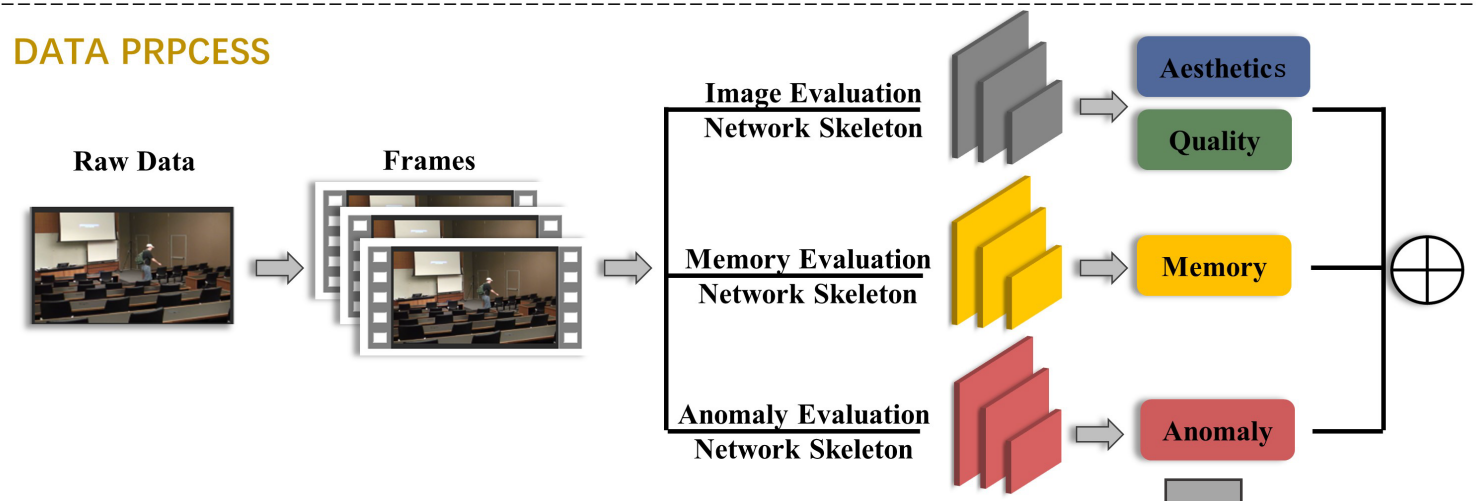
T2.2. Feature-Level

T2.3. Frame-Level

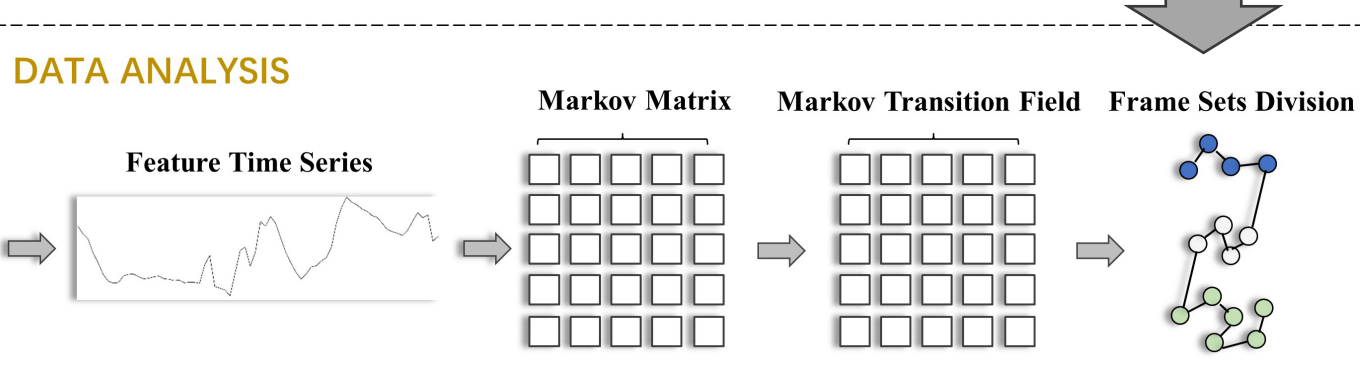


SurVizor: Workflow

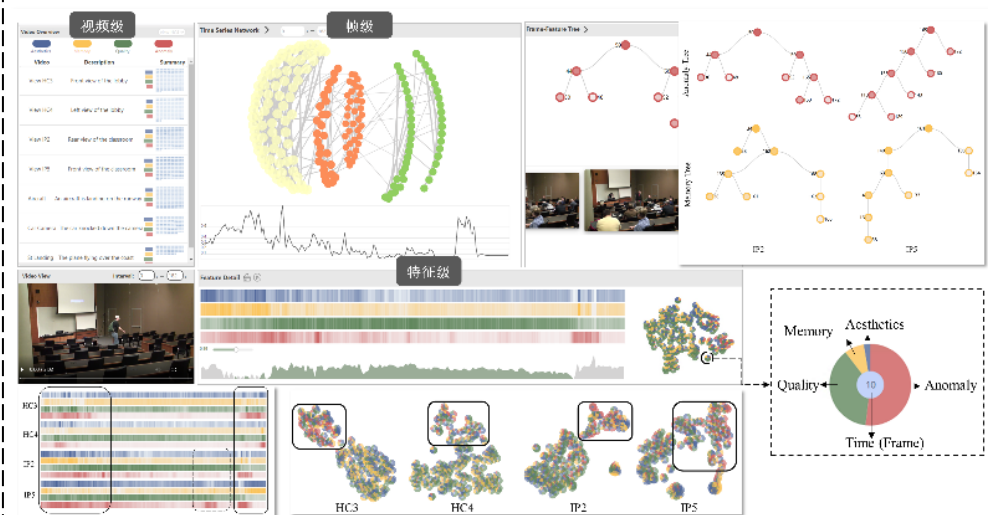
DATA PROCESS



DATA ANALYSIS



VISUAL ANALYSIS



Video-level Visual Design and Analysis

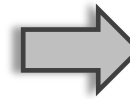
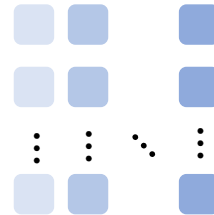
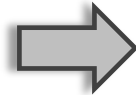
Quantitative Characterization

Visual Encoding

Interaction Analysis

frame $F = \text{fused feature } f$

video $V = [f_{aes}, f_q, f_m, f_{ano}]$



Video Overview View-HC4

Aesthetics Memory Quality Anomaly

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	

Feature-level Visual Design and Analysis

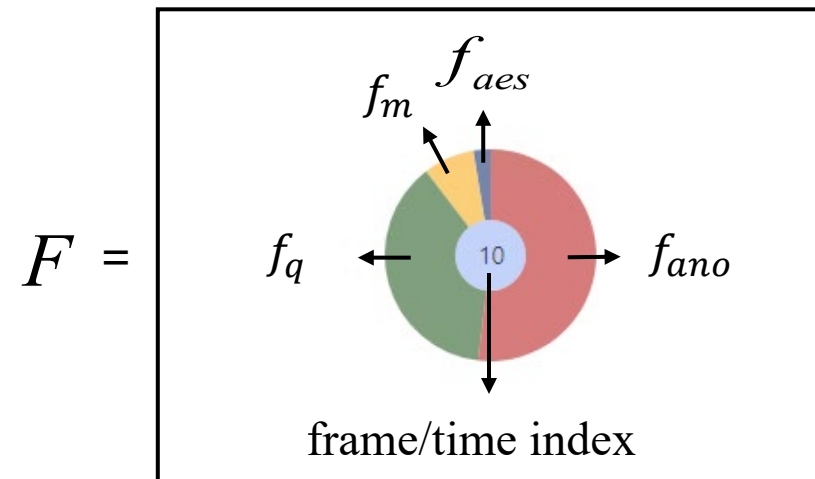
□ Frames Sequence



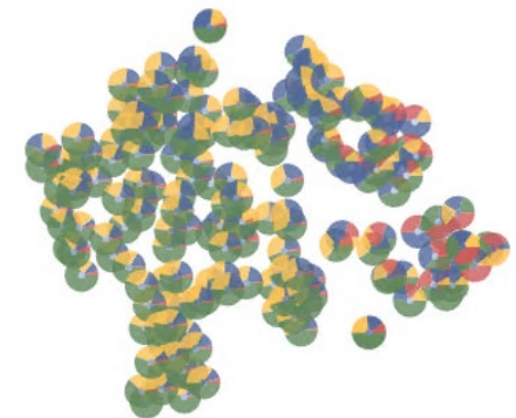
□ Vector Characterization

$$\text{frame } F = V [f_{aes}, f_q, f_m, f_{ano}]$$

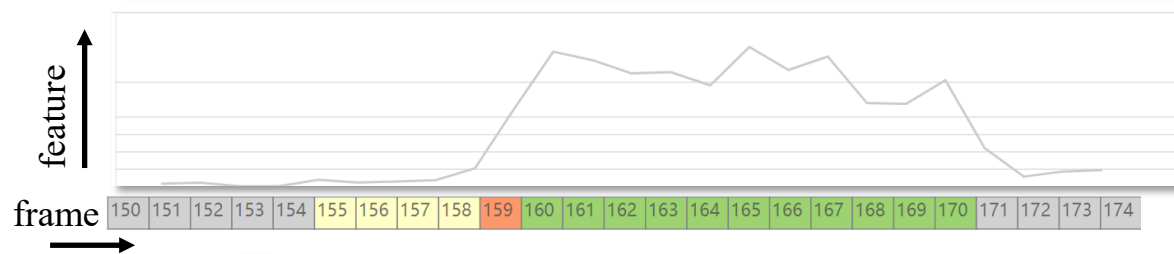
□ Visual Encoding



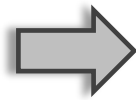
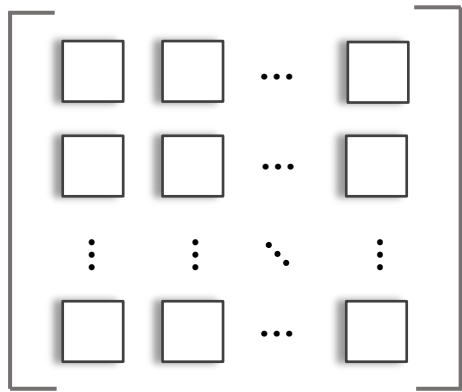
□ Projection



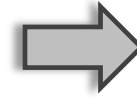
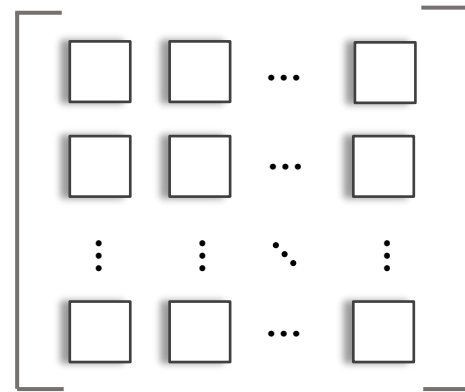
Frame-level Visual Design and Analysis



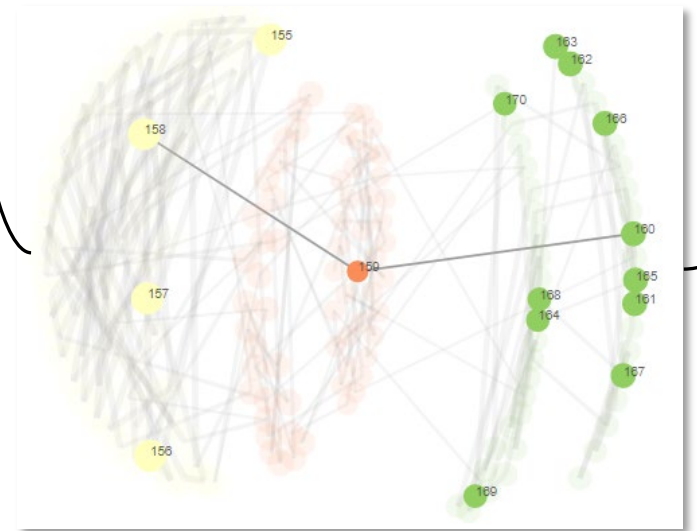
□ Markov Matrix



□ Markov Transition Field



□ Visual Encoding



Normal Transition: 155 → 156 → 157

Abnormal Transition: 158 → 159 → 160

SurVizor: Case Demonstration

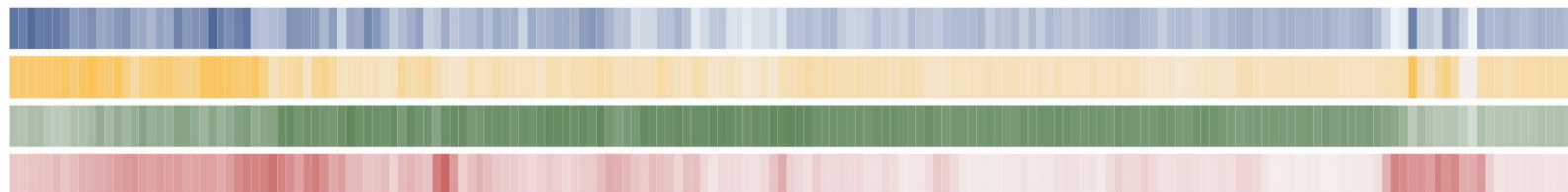
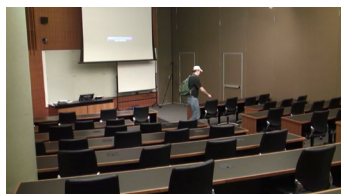
- **Location: Campus - Classroom**
- **Duration: 5mins**



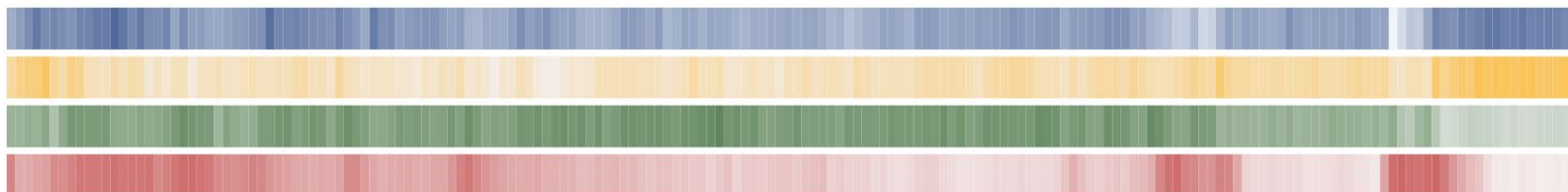
SurVizor: Case Demonstration

Feature Detail View (Feature Level)

IP2



IP5



PS.  Aesthetics  Memory  Quality  Anomaly

SurVizor: Case Demonstration

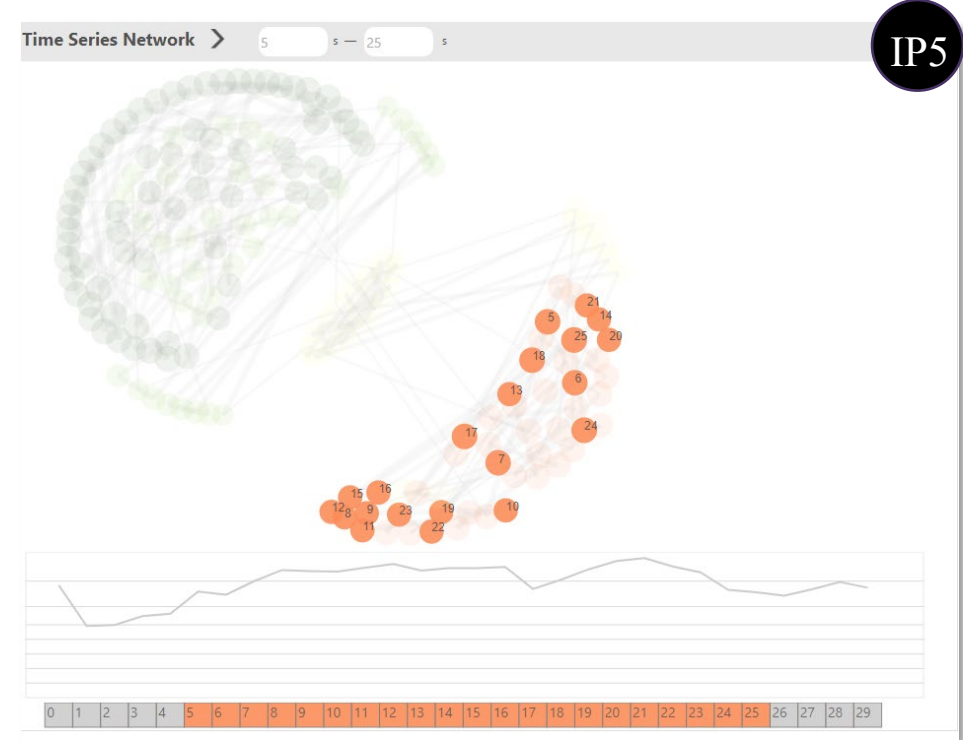
Time Series Network – Anomaly (Frame Level) 5s – 25s



5



20



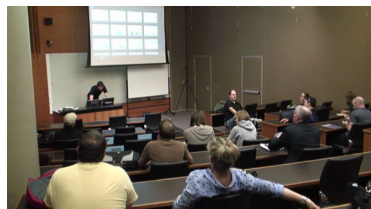
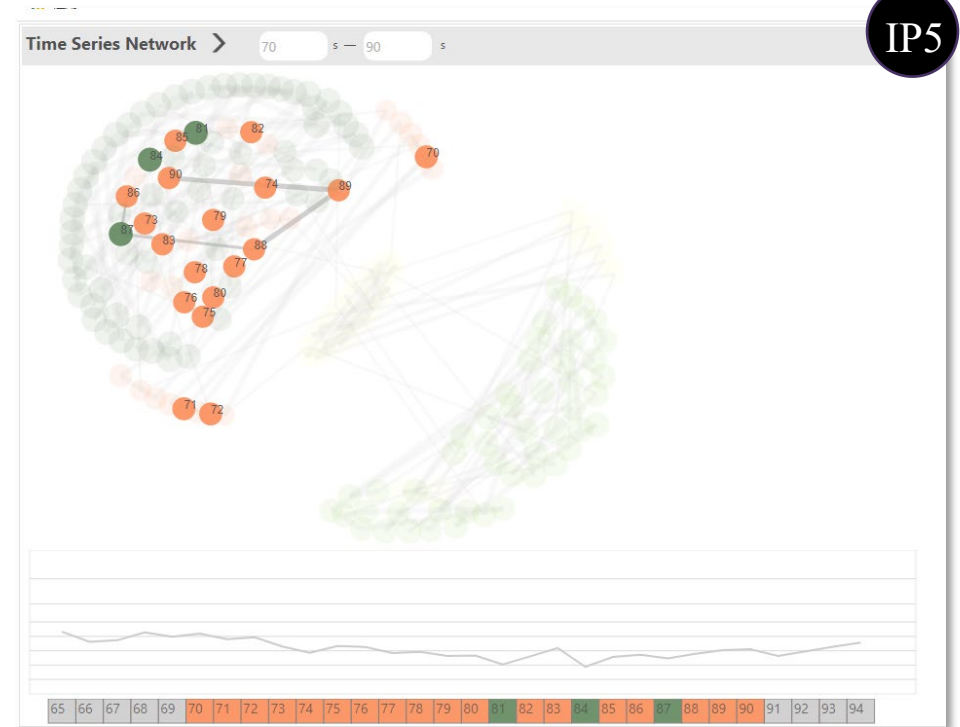
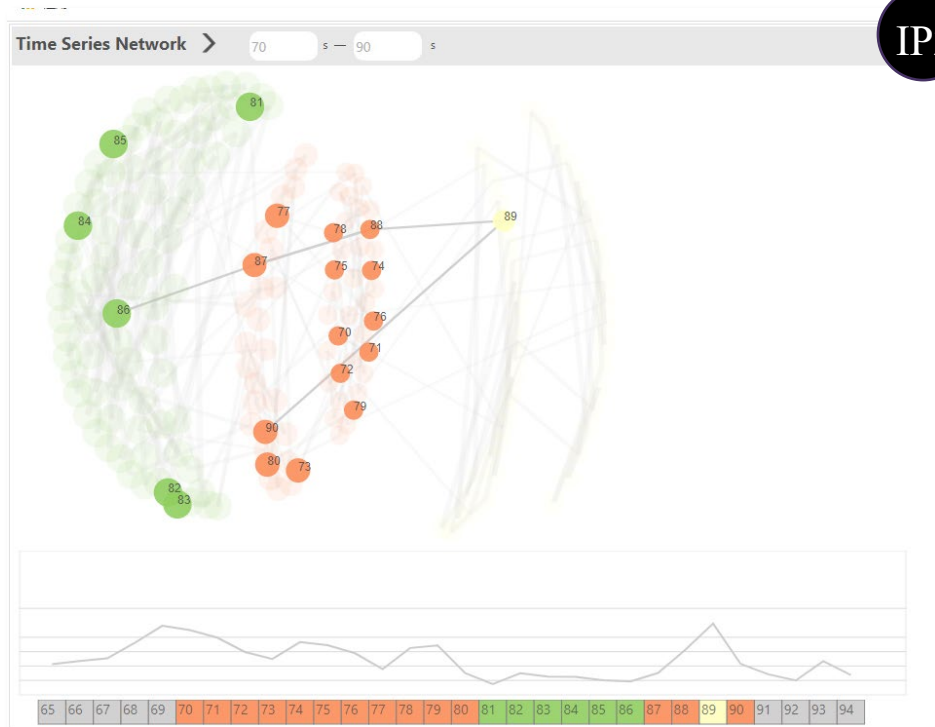
5



20

SurVizor: Case Demonstration

Time Series Network – Anomaly (Frame Level) 70s – 90s



75



85



89



75



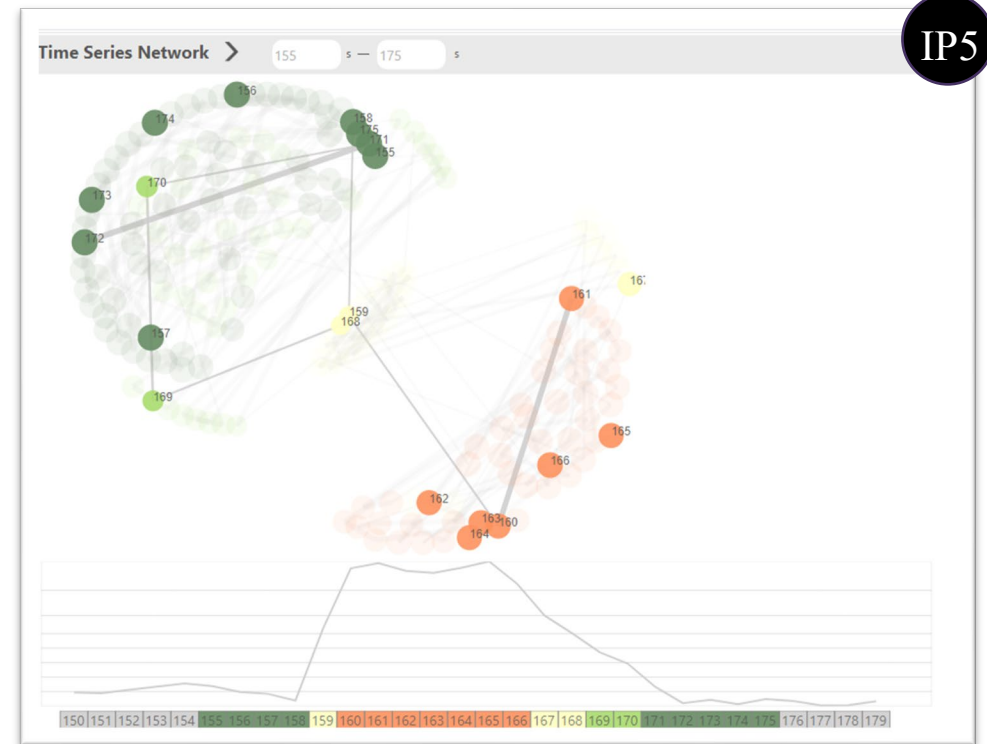
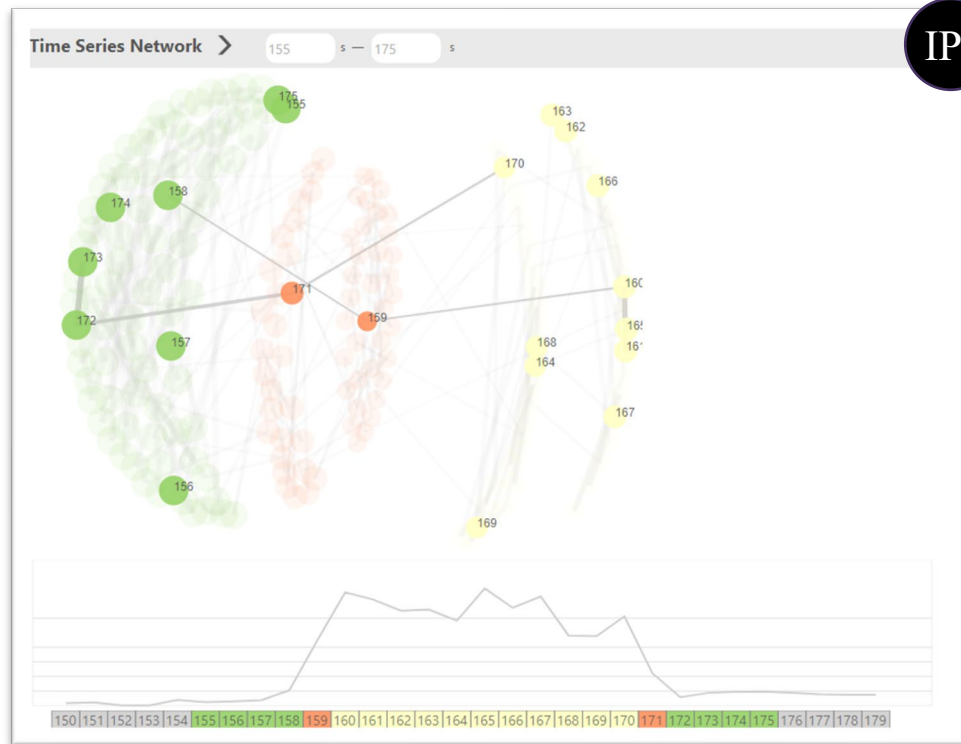
85



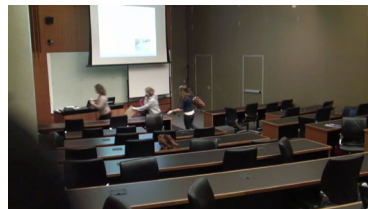
89

SurVizor: Case Demonstration

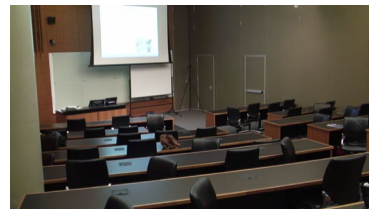
Time Series Network – Anomaly (Frame Level) 155s – 175s



155



165



175



155



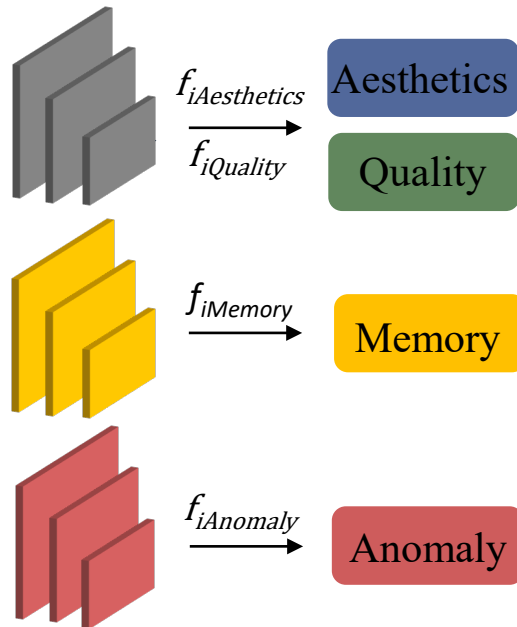
165



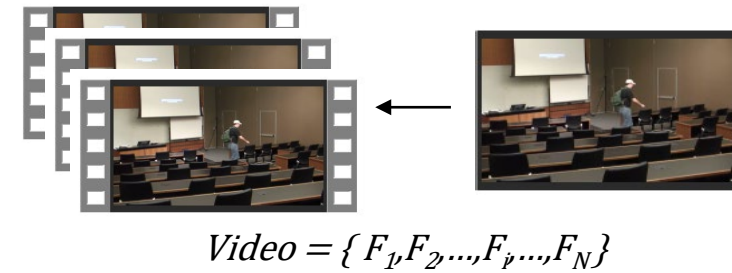
175

SurVizor: Discussion

Feature Selection



Dynamic Sampling



References

- [1] L. Guo, L. Zou, et al. "**VATLD**: A Visual Analytics System to Assess, Understand and Improve Traffic Light Detection." IEEE Transactions on Visualization and Computer Graphics. 2020.
- [2] T. Tang, Y. Wu, et al. "**VideoModerator**: A Risk-aware Framework for Multimodal Video Moderation in E-Commerce." IEEE Transactions on Visualization and Computer Graphics. 2021.
- [3] H. Zeng, X. Wang, et al. "**EmoCo**: Visual Analysis of Emotion Coherence in Presentation Videos." IEEE Transactions on Visualization and Computer Graphics. 2019.
- [4] H. Zeng, X. Shu, et al. "**EmotionCues**: Emotion-oriented Visual Summarization of Classroom Videos." IEEE Transactions on Visualization and Computer Graphics. 2020.
- [5] A. Wu and H. Qu. "Multimodal Analysis of Video Collections: Visual Exploration of **Presentation Techniques** in TED Talks." IEEE Transactions on Visualization and Computer Graphics. 2018.
- [6] H. Li, M. Xu, et al. "A Visual Analytics Approach to Facilitate the Proctoring of **Online Exams**." Proceedings of CHI Conference on Human Factors in Computing Systems. 2021.