Video Semantic Analytics and Visualization

Tong Li

TongLi97.github.io

ZJUTVIS Lab

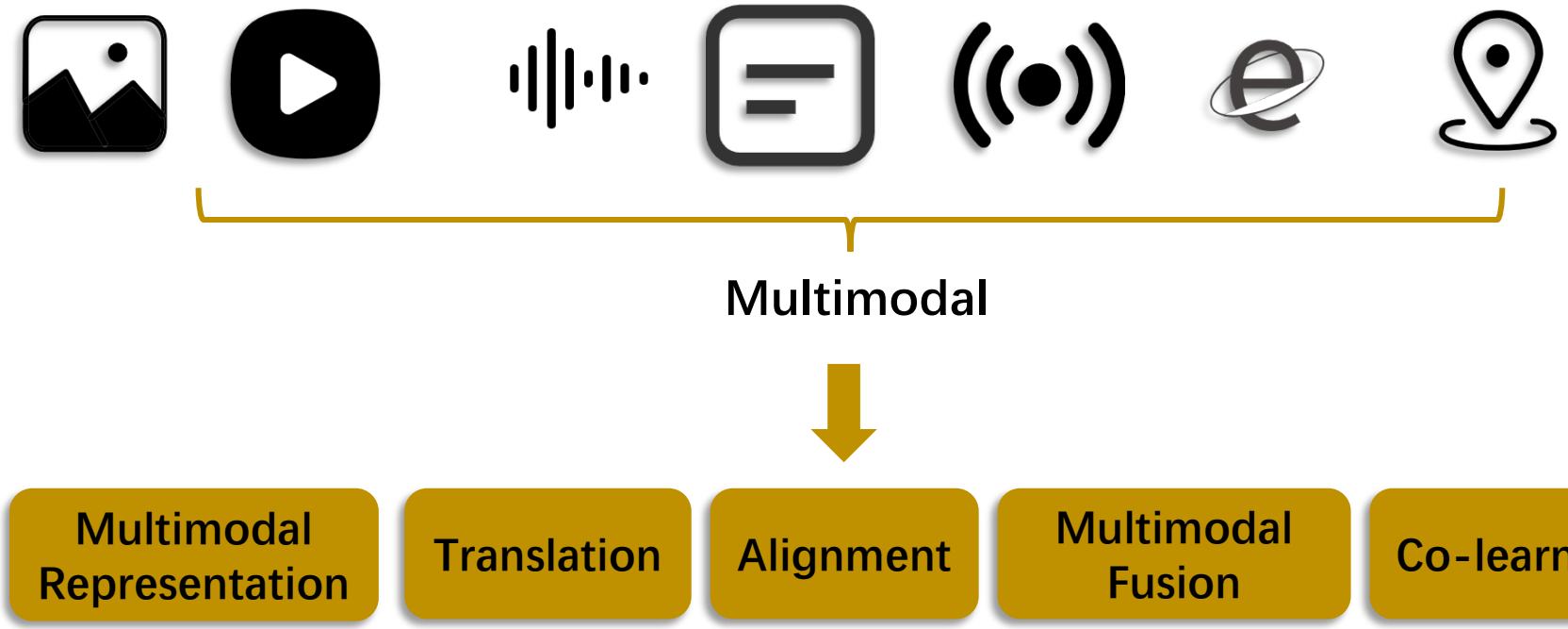
Zhejiang University of Technology

Outline

- Background
  - Multimedia Data
  - Video Data
  - Visual Analytics
- Related Papers
  - Media Video Vis
  - Entertainment Video Vis
  - Sport Video Vis
  - Medical Video Vis
  - Surveillance Video Vis
  - Summary
- Surveillance Video
  - Summary
  - Goals and Challenges
  - Video Understanding

Multimedia Data

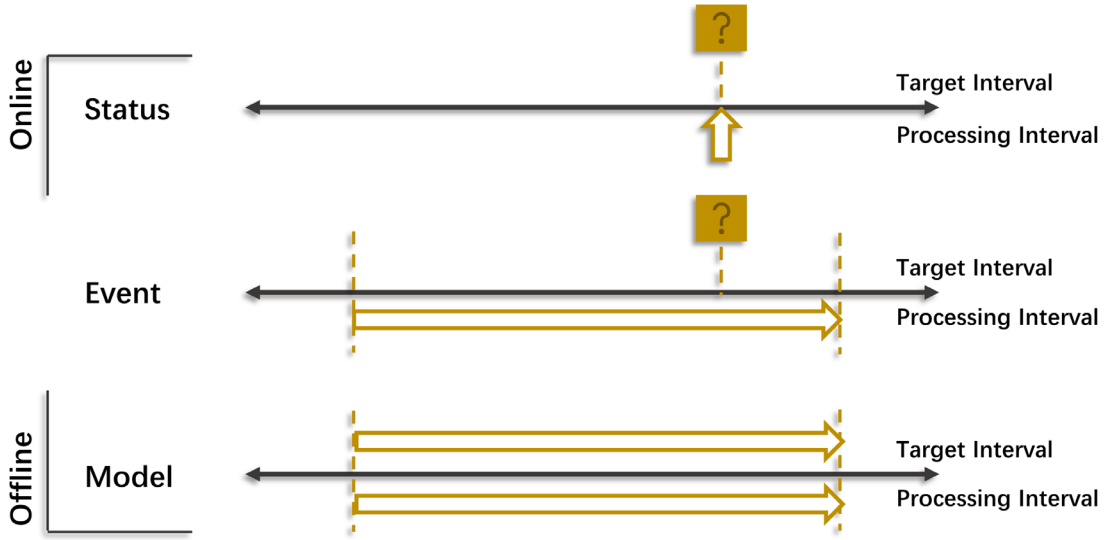
- Visual data
- Audio data
- Text data
- Sensor data
- Other data



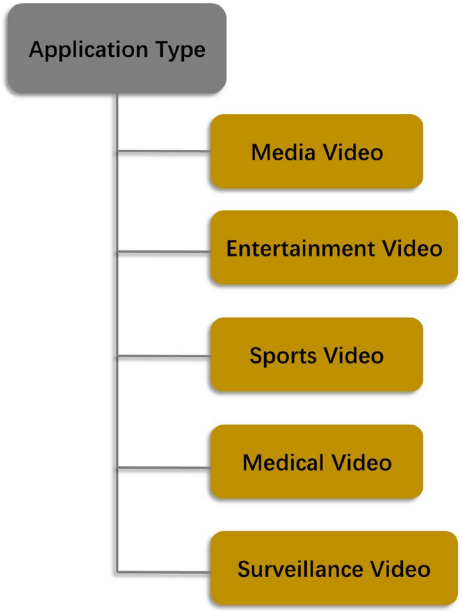
Video Data

Input State

Online || Offline

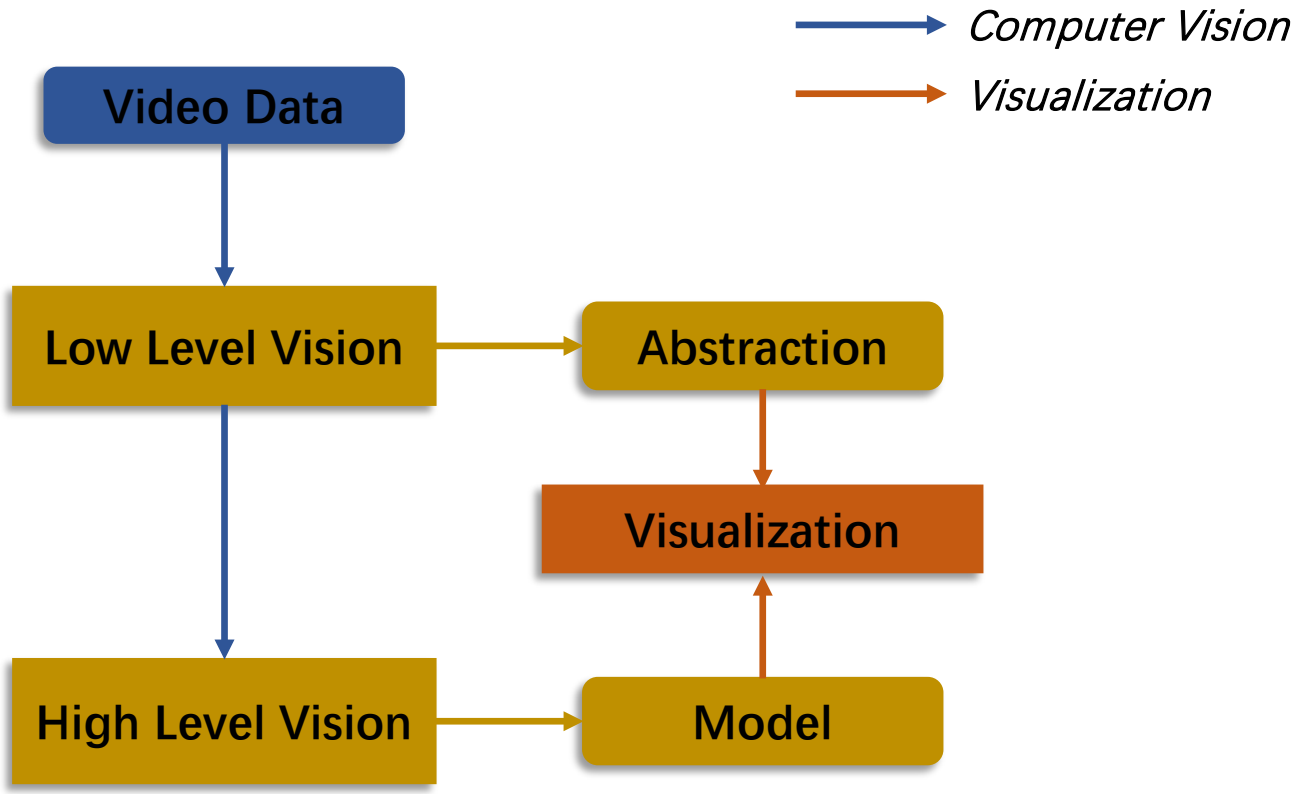


Application Type



Video Analytics

- Low Level Vision
  - Optical Flow Estimation
  - Image Segmentation
  - Feature Extraction
  - ……
- High Level Vision
  - Detection, Recognition, Tracking ……

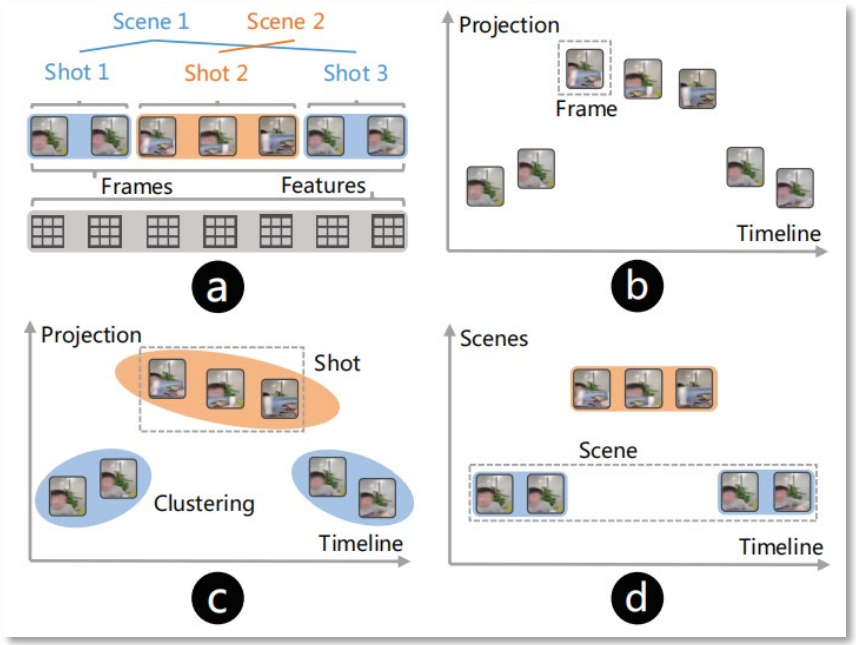


Outline

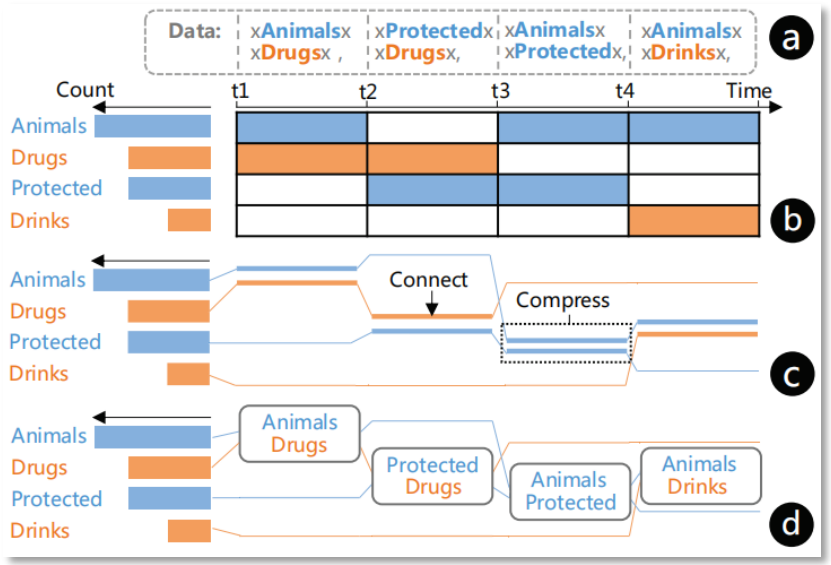
- Background
  - Multimedia Data
  - Video Data
  - Visual Analytics
- Related Papers
  - Media Video Vis
  - Entertainment Video Vis
  - Sport Video Vis
  - Medical Video Vis
  - Surveillance Video Vis
  - Summary
- Surveillance Video
  - Summary
  - Goals and Challenges
  - Video Understanding

Media Video Vis

- **Purpose**
  Risk assessments on e-commerce videos.
- **Target User**
  Video Reviewer
- **Data**
  E-commerce Video Data: Visual and Audio



(1) Video Frame



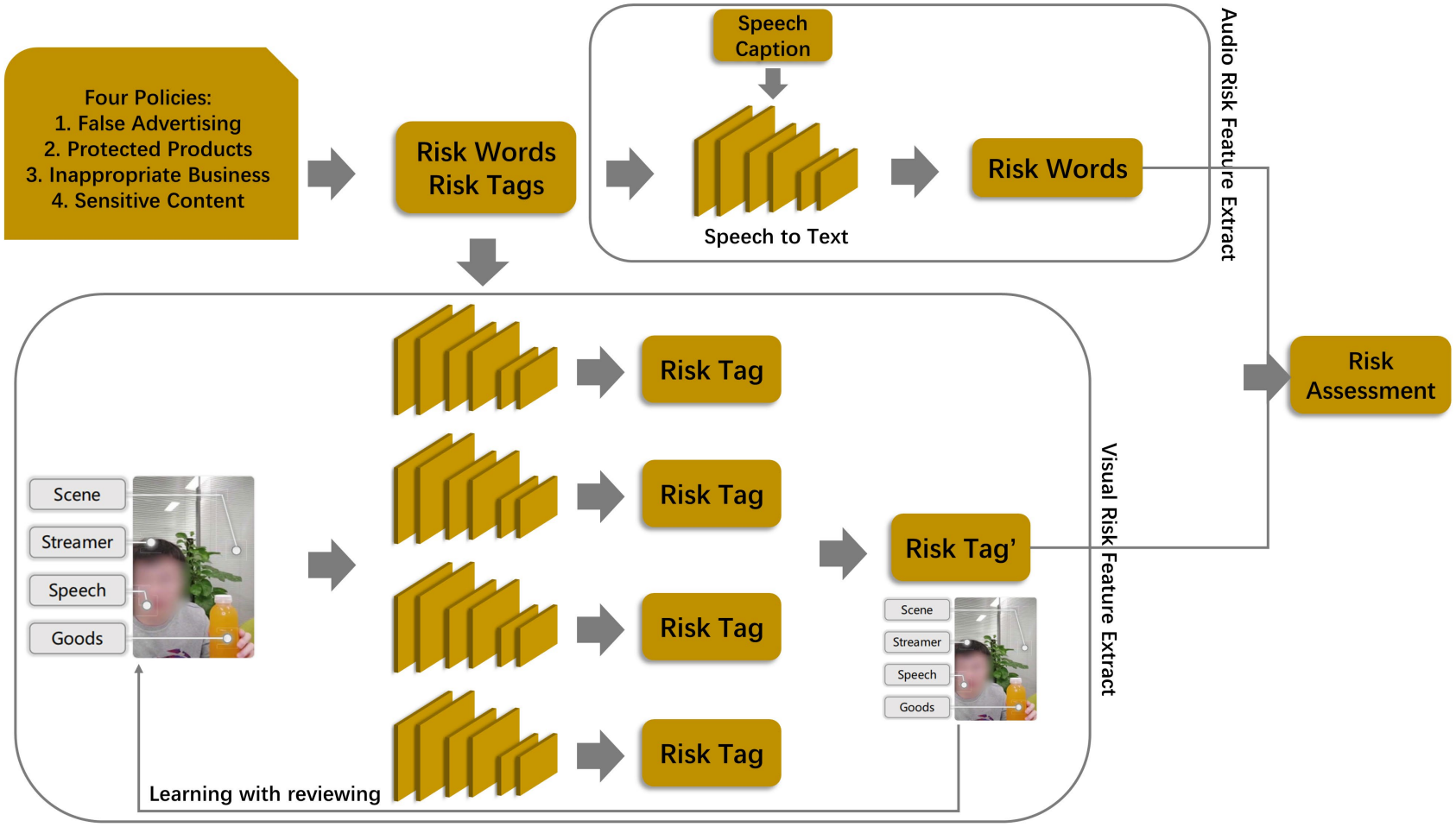
(2) Audio Content

(3)

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.

Media Video Vis

- Solution
- Visualization



(1) Overview

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.

Media Video Vis

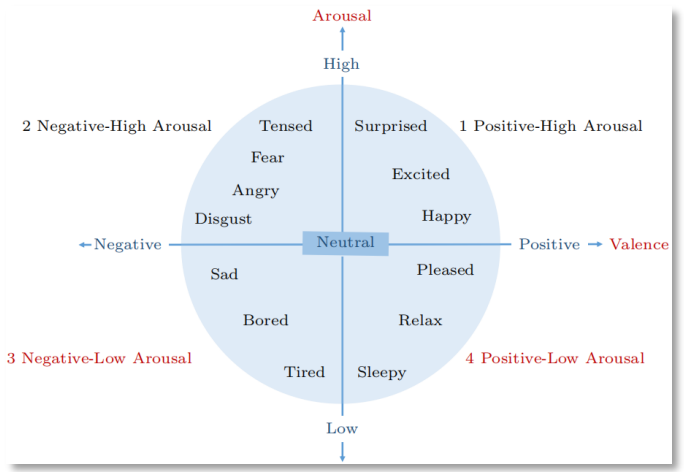
Cons

There is a lack of detailed descriptions of Risk Tags and Risk Words.
The accuracy of model is not mentioned.
It would be better to draw a pipeline for data processing.
Waste of pixel space.

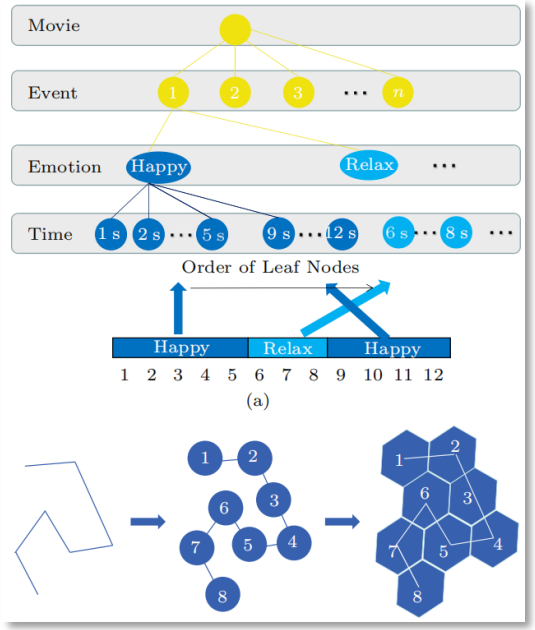
T. Tang, Y. Wu, et al. “VideoModerator: A Risk-aware Framework for Multimodal Video Moderation in E-Commerce.” IEEE Transactions on Visualization and Computer Graphics, 2021.

Entertainment Video Vis I

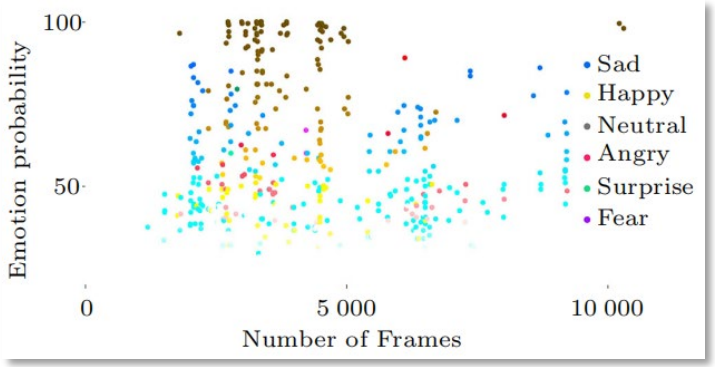
- **Purpose**
  Explore, understand, and search movie content through the angle of emotion.
- **Target User**
  Audience and Editor
- **Data**
  Users' assessments of movies - Subjective
  Characters' facial expression based on deep learning - Objective
- **Solution**
- **Visualization**



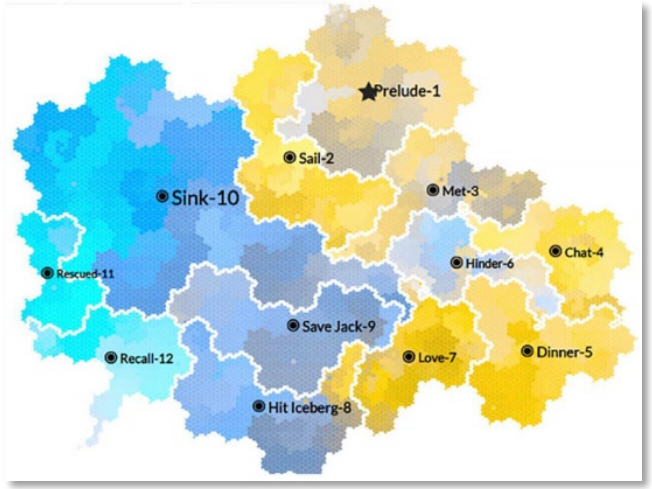
(1) Emotional Model



(2) Emotion Map



(4) Characters Emotion



(3) Event Map

C. Ma, et al. "EmotionMap: Visual Analysis of Video Emotional Content on a Map." Journal of Computer Science and Technology, 2020.

Entertainment Video Vis I

Cons

The preliminary evaluation work was not described in detail.
The method of dividing the video into events is not mentioned.
There is no correlation between the two kinds of sentiment data.

C. Ma, et al. "EmotionMap: Visual Analysis of Video Emotional Content on a Map." Journal of Computer Science and Technology, 2020.

Entertainment Video Vis II

- **Purpose**

  Analyze key factors of an inspirational speech and quantitatively evaluate the effectiveness of the factors.

- **Target User**

  Speakers and Speech Experts

- **Data**

  Speech Video, Script, Metadata, Information(Region, Year, Level …)

  Feature Emotional Data(Facial, Text, Audio) Non-emotional Data

  Factors List

(1)

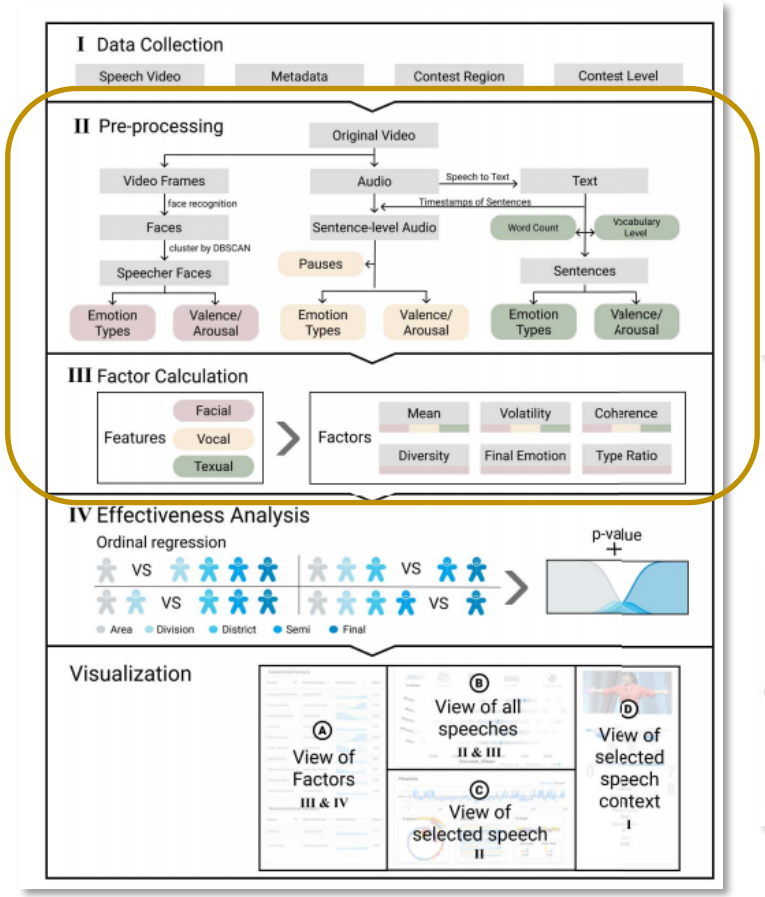
Factor	Modality	Type(p-value)	Type(p-value)
Average	Facial	Arousal(0.006*)	Valence(0.431)
	Textual	Arousal(0.215)	Valence(0.088)
	Vocal	Arousal(0.016*)	Valence(0.017*)
Volatility	Facial	Arousal(0.020*)	Valence(0.006*)
	Textual	Arousal(0.433)	Valence(0.438)
	Vocal	Arousal(0.235)	Valence(0.845)
Diversity	Facial	Across Emotion Type(0.120)	
Final	Facial	Arousal(0.002*)	Valence(0.020*)
Coherence	All	Arousal(0.124)	Valence(0.051)
Ratio	Facial	Happy(0.001*)	Sad(0.0736)
		Fear(0.582)	Angry(0.292)
		Surprise(0.115)	Disgust(0.306)
		Neutral(0.488)	-
Pauses	Vocal	Pauses(0.271)	-
Vocabulary	Textual	Vocabulary(0.089)	-

(2) Factors List

K. Maher, et al. "E-ffective: A Visual Analytic System for Exploring the Emotion and Effectiveness of Inspirational Speeches." IEEE Transactions on Visualization and Computer Graphics, 2021.

Entertainment Video Vis II

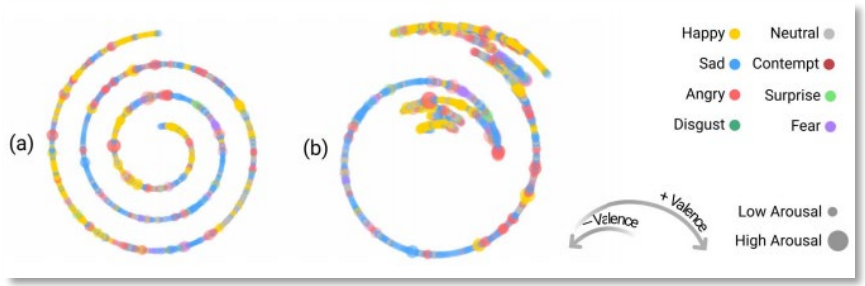
- Solution
- Visualization



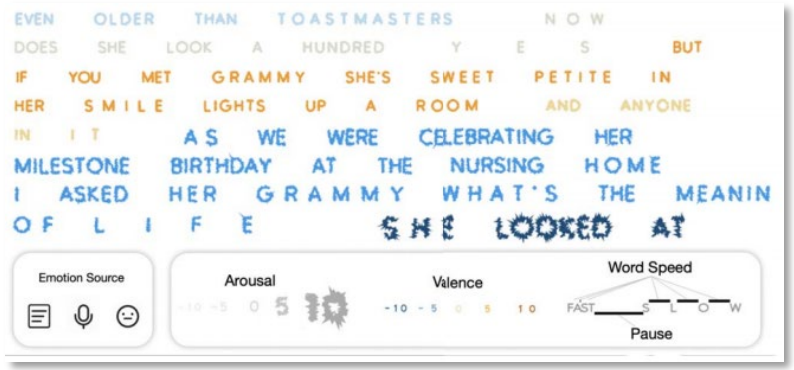
(1) Overview

Module	Description
E-factor	To evaluate hypotheses of interest about speech factors using the cumulative data of all speeches.
E-type	To understand discrete emotional data contained in emotional types, as well as their distribution over time.
E-script	To understand the emotion in speech scripts.
E-spiral	To provide an intuitive way of understanding the emotional shifts within speeches.
E-similarity	To understand the similarity and the effectiveness estimation of speech factors in speeches.
E-distribution	To understand distribution of factor effectiveness among speech levels.

(2) Visualization Module



(3) E-spiral



(4) E-script

K. Maher, et al. "E-ffective: A Visual Analytic System for Exploring the Emotion and Effectiveness of Inspirational Speeches." IEEE Transactions on Visualization and Computer Graphics, 2021.

Entertainment Video Vis II

Cons

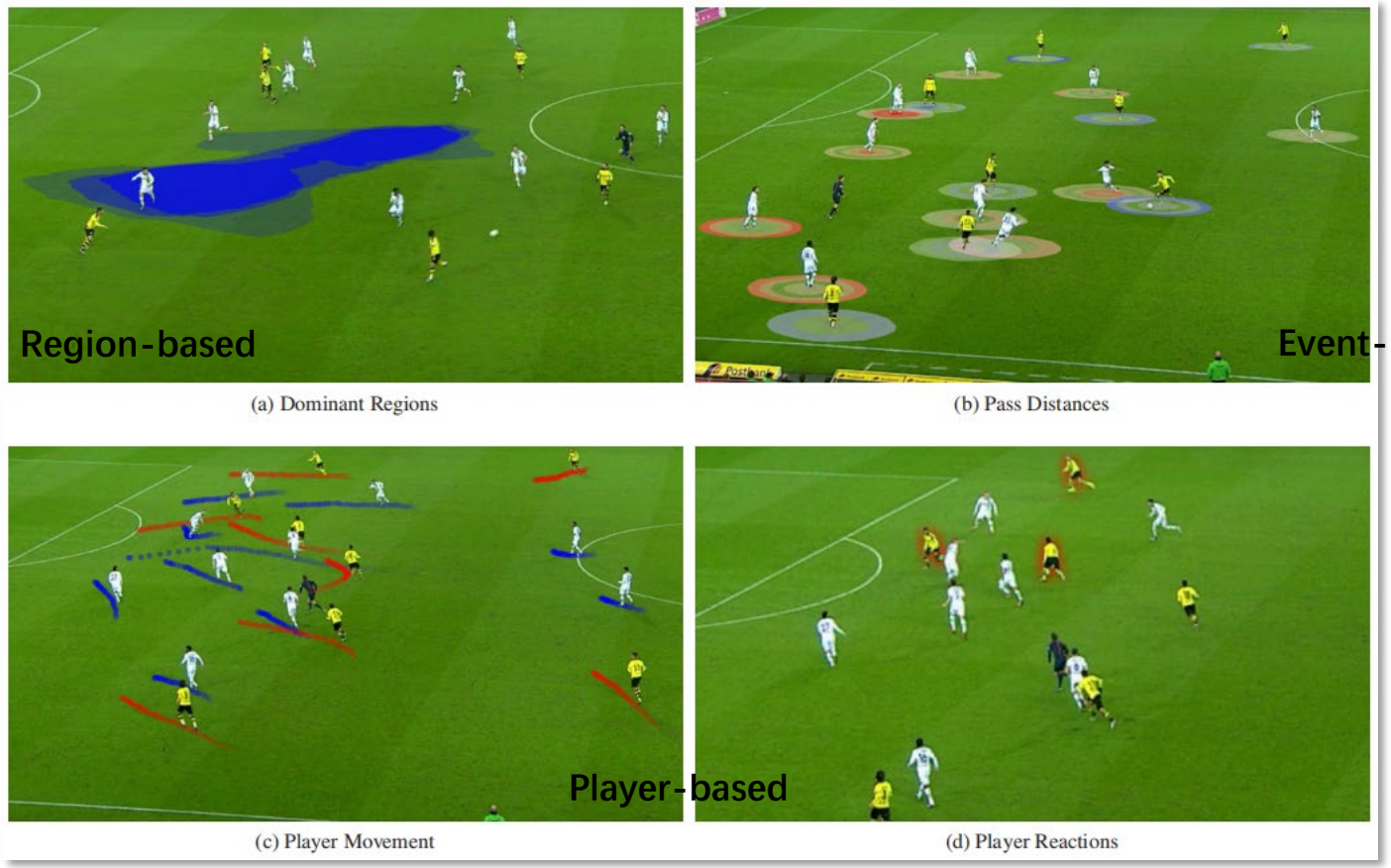
The definition of *Valence* and *Arousal* is not explained.

The accuracy of model is not mentioned.

K. Maher, et al. “E-ffective: A Visual Analytic System for Exploring the Emotion and Effectiveness of Inspirational Speeches.” IEEE Transactions on Visualization and Computer Graphics, 2021.

Sport Video Vis

- **Purpose**
  Analyze soccer videos to help analysts gain insights into player behavior and team tactics.
- **Target User**
  Team Sport Analysts
- **Data**
  Soccer Match Video



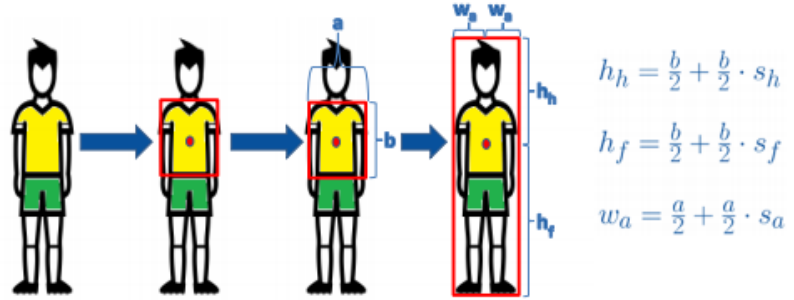
Event-based

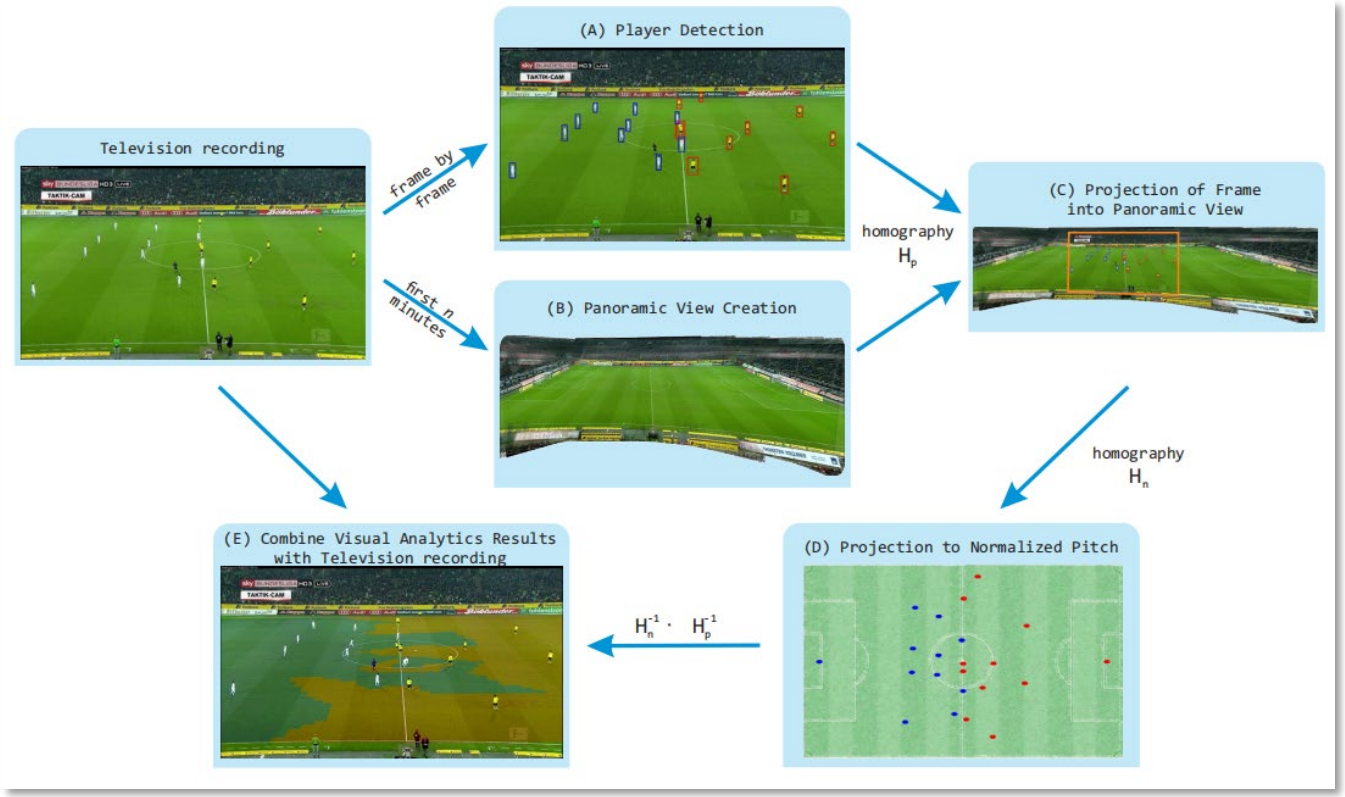
(1) Visual Analysis

M. Stein, H. Janetzko, et al. “Bring it to the pitch: Combining Video and Movement Data to Enhance Team Sport Analysis.” IEEE Transactions on Visualization and Computer Graphics, 2017.

Sport Video Vis

- Solution
  - Player Detection, Ball Detection
  - Player Trajectory
- Visualization





(1) Overview

M. Stein, H. Janetzko, et al. “Bring it to the pitch: Combining Video and Movement Data to Enhance Team Sport Analysis.” IEEE Transactions on Visualization and Computer Graphics, 2017.

Sport Video Vis

- Cons

  The tasks of domain experts are not rich enough.

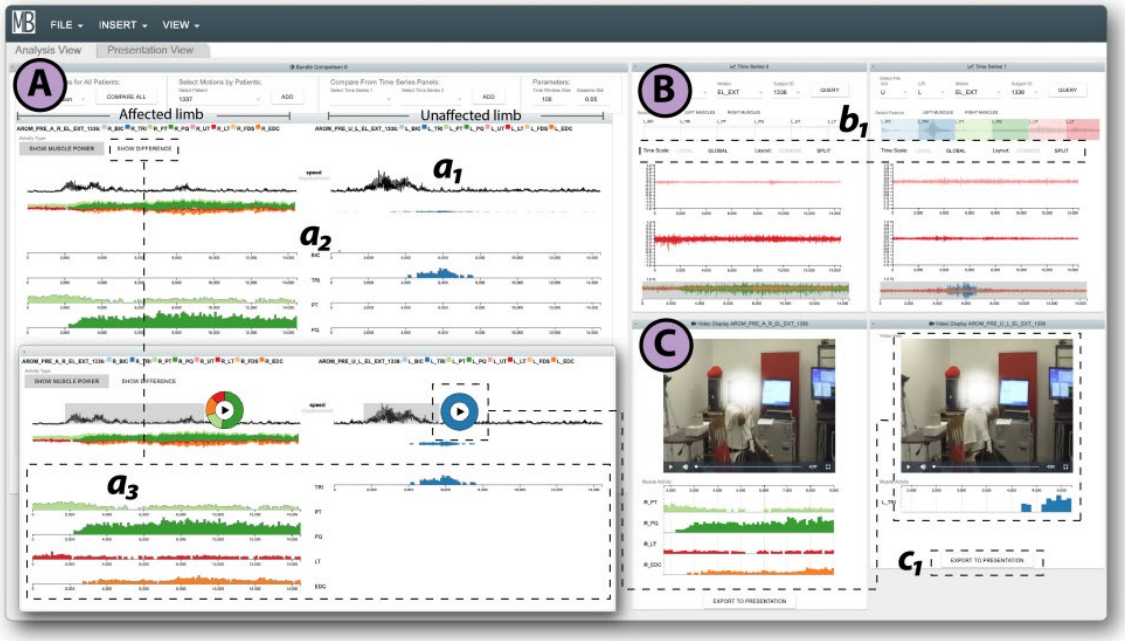
  Event-based analysis is relatively simple.

  Color design conflicts.

M. Stein, H. Janetzko, et al. “Bring it to the pitch: Combining Video and Movement Data to Enhance Team Sport Analysis.” IEEE Transactions on Visualization and Computer Graphics, 2017.

Medical Video Vis

- **Purpose**
  Study the muscle activity patterns of patients with brachial plexus injuries.
- **Target User**
  Doctor
- **Data**
  Muscle Signals, Motion Data, Video Record
- **Solution**
- **Visualization**

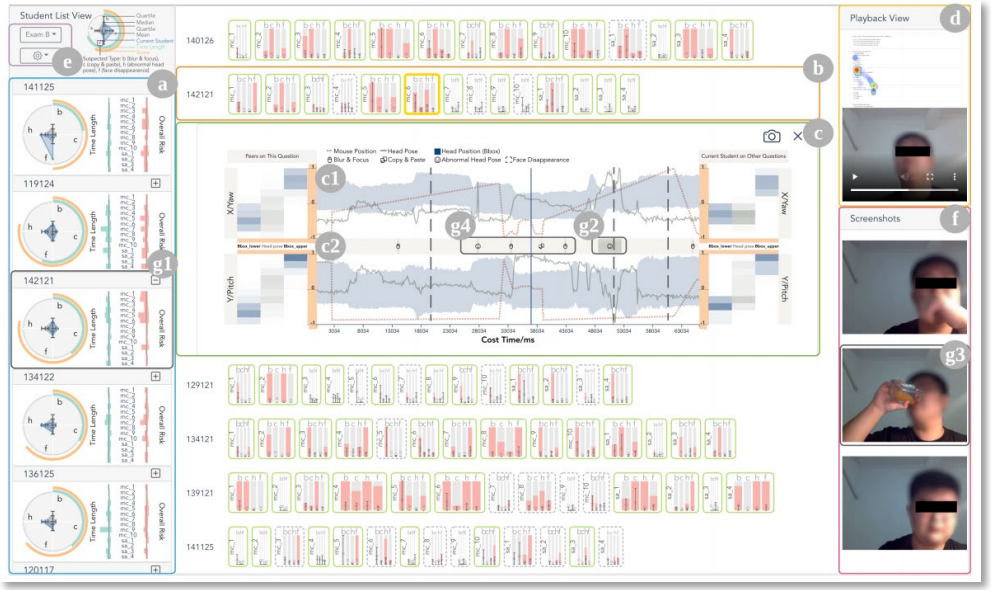


(1)

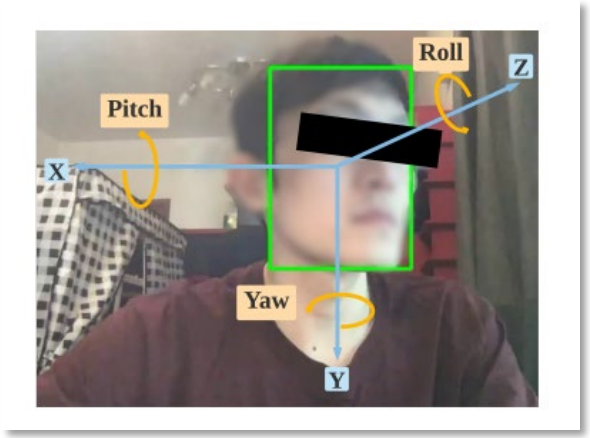
G. Chan, L.G. Nonato, et al. "Motion Browser: Visualizing and Understanding Complex Upper Limb Movement under Obstetrical Brachial Plexus Injuries." IEEE Transactions on Visualization and Computer Graphics, 2019.

Surveillance Video Vis

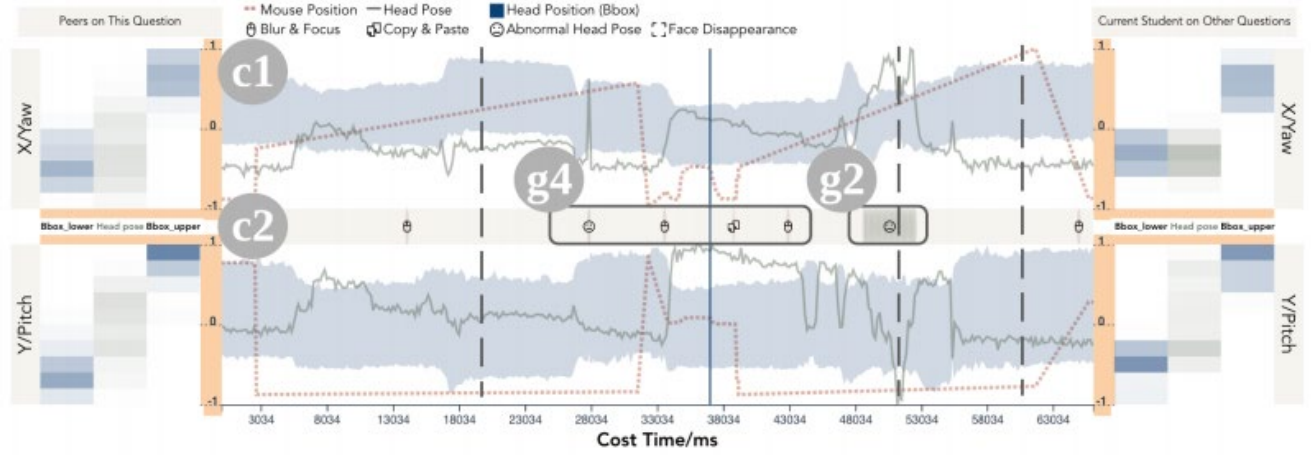
- **Purpose**
  Analysis of cheating behavior in online exams.
- **Target User**
  Teacher
- **Data**
  Mock Online Exam
  Cheating Types: Local Environment, Computer
  Webcam Video Data → Abnormal Head Movement
  Mouse Movement (JavaScript Plugin) → Abnormal Mouse Movement



(3)



(1) Head Pose

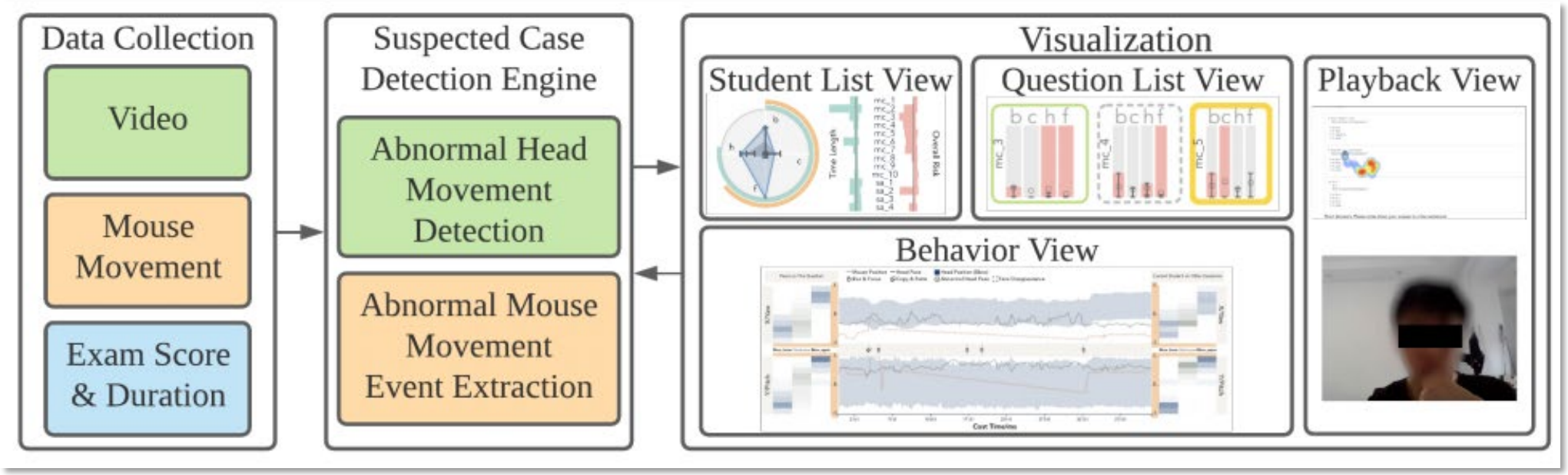


(2) Mouse and Head Movement

H. Li, M. Xu, et al. "A Visual Analytics Approach to Facilitate the Proctoring of Online Exams."
Proceedings of the CHI Conference on Human Factors in Computing Systems. 2021.

Surveillance Video Vis

- Solution
  - Abnormal Head Movement Detection: Face Disappearance, Abnormal Head Pose
  - Abnormal Mouse Movement Detection: Blur, Focus, Copy, Paste, Mousemove, Mousewheel
  - Overall Risk Estimation
- Visualization
- Cons



(1) Overview

H. Li, M. Xu, et al. "A Visual Analytics Approach to Facilitate the Proctoring of Online Exams." Proceedings of the CHI Conference on Human Factors in Computing Systems. 2021.

Summary

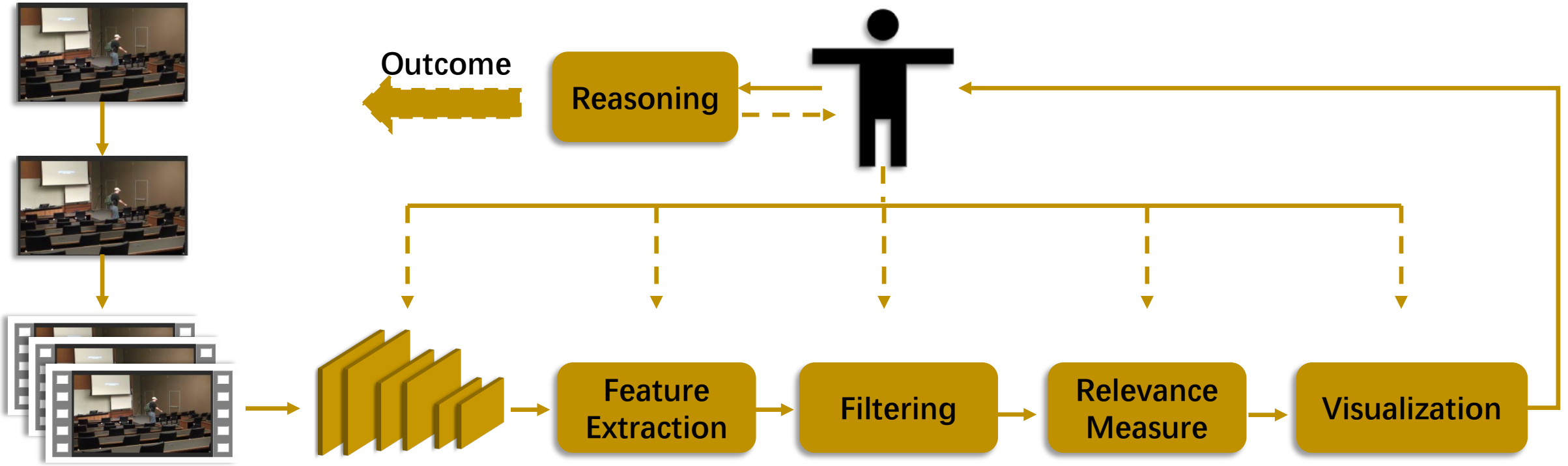
	Media		Entertainment				Sport		Medical	Surveillance		
Info Meta	John et al.	Tang et al.	Ma et al.	Wu et al.	Zeng et al.	Maher et al.	Stein et al.	Chen et al.	Chan et al.	Lee et al.	Zeng et al.	Li et al.
	2019	2021	2020	2018	2019	2021	2017	2021	2019	2019	2020	2021
	SCI IV	TVCG	SCI II	TVCG	TVCG	TVCG	TVCG	TVCG	TVCG	TVCG	TVCG	CHI
	1	0	2	12	17	0	90	3	5	29	18	5
Data-Source	News	E-commerce	Movie	TED	TED	Speech Contest	Soccer Match	Table tennis Match	Video	Traffic Video	Classroom Video	Examination Video
Data-Multimodal												
Model-Usage												
Model-Accuracy												
Research Focus												

Outline

- Background
  - Multimedia Data
  - Video Data
  - Visual Analytics
- Related Papers
  - Media Video Vis
  - Entertainment Video Vis
  - Sport Video Vis
  - Medical Video Vis
  - Surveillance Video Vis
  - Summary
- Surveillance Video
  - Summary
  - Goals and Challenges
  - Video Understanding

Summary

- Manual Inspection: Labor-intensive Tasks
- Machine Intelligence: Inaccurate Results



Surveillance Video

Analytics Target

Reduce the time of watching videos.
Understand video with low cost.

Data Challenge

Big Data、Uneven Quality
Noise Data
Loose Structures or Without Story Units

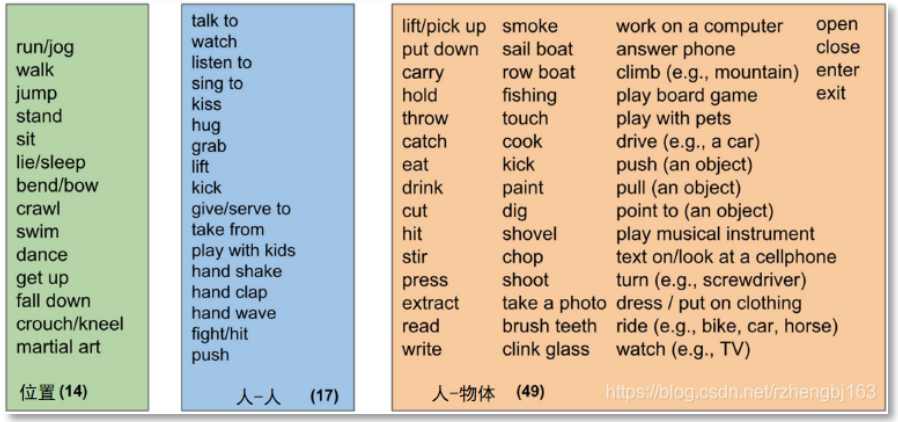
Visualization Challenge

Limited pixel space.

Video Understanding based on Action Recognition

Action Recognition

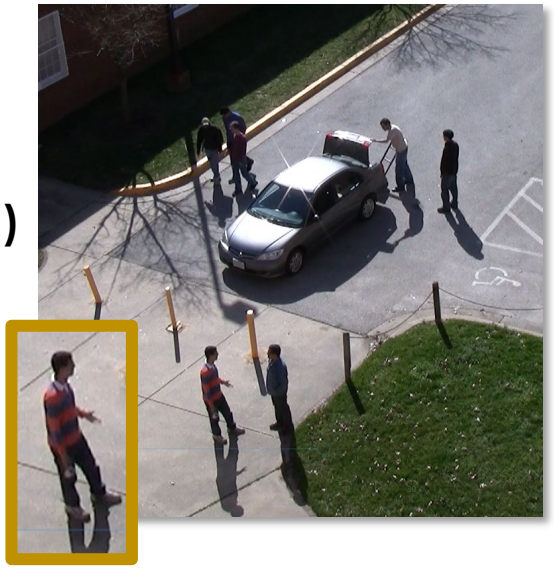
It is difficult to precisely define the boundary and length of the action.
The accuracy is difficult to reach 100%.
It is not possible to label all human actions.



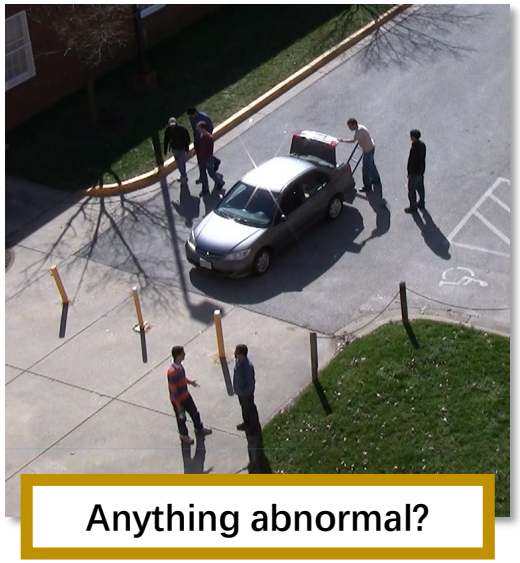
AVA Actions Dataset

Event Understanding

Normal or abnormal / Key or Universal (Vaguely Define)
"Key events" accounted for a relatively low proportion.



a. Clearly Define



b. Vaguely Define

References

[1] M. John, K. Kurzhals and T. Ertl. "Visual Exploration of Topics in Multimedia News Corpora." Proceedings of International Conference Information Visualization. 2019.
[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] C. Ma, J. Song, et al. "EmotionMap: Visual Analysis of Video Emotional Content on a Map." Journal of Computer Science and Technology. 2020.
[4] 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.
[5] H. Zeng, X. Wang, et al. "EmoCo: Visual Analysis of Emotion Coherence in Presentation Videos." IEEE Transactions on Visualization and Computer Graphics. 2019.
[6] K. Maher, Z. Huang, et al. "E-ffective: A Visual Analytic System for Exploring the Emotion and Effectiveness of Inspirational Speeches." IEEE Transactions on Visualization and Computer Graphics. 2021.
[7] M. Stein, H. Janetzko, et al. "Bring it to the pitch: Combining Video and Movement Data to Enhance Team Sport Analysis." IEEE Transactions on Visualization and Computer Graphics. 2017.
[8] Z. Chen, S. Ye, et al. "Augmenting Sports Videos with VisCommentator." IEEE Transactions on Visualization and Computer Graphics. 2021.
[9] G. Chan, L.G. Nonato, et al. "Motion Browser: Visualizing and Understanding Complex Upper Limb Movement under Obstetrical Brachial Plexus Injuries." IEEE Transactions on Visualization and Computer Graphics. 2019.
[10] C. Lee, Y. Kim, et al. "A Visual Analytics System for Exploring, Monitoring, and Forecasting Road Traffic Congestion." IEEE Transactions on Visualization and Computer Graphics. 2019.
[11] H. Zeng, X. Shu, et al. "EmotionCues: Emotion-oriented Visual Summarization of Classroom Videos." IEEE Transactions on Visualization and Computer Graphics. 2020.
[12] 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.