Video Semantic Analytics and Visualization



Tong Li

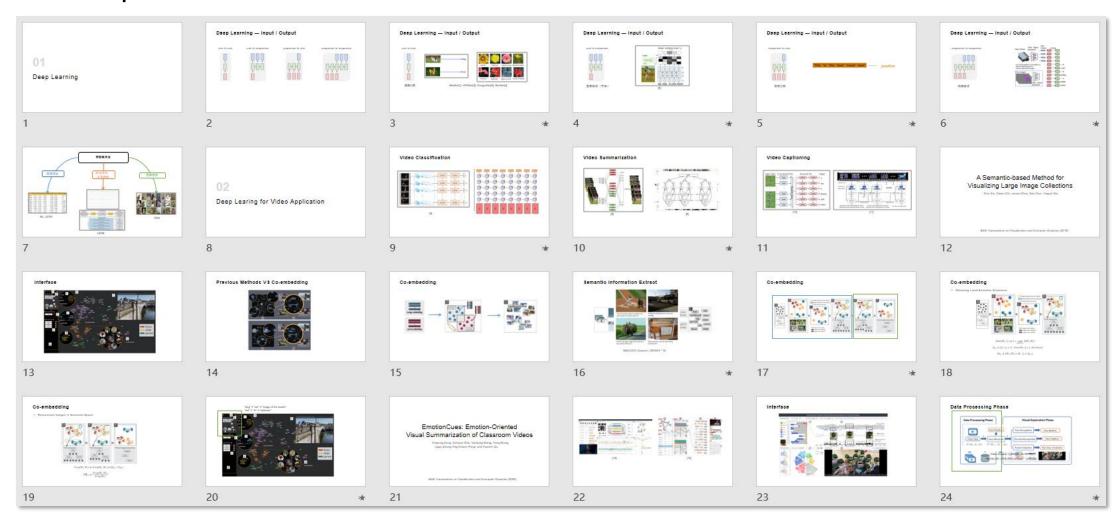


ZJUTVIS Lab
Zhejiang University of Technology

Previous Report

□ Deep Learning for Video Application

□ Related Paper

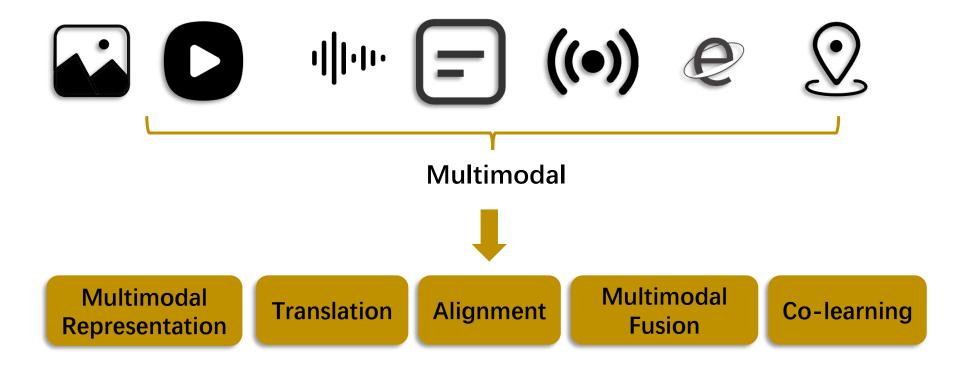


Outline

- 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

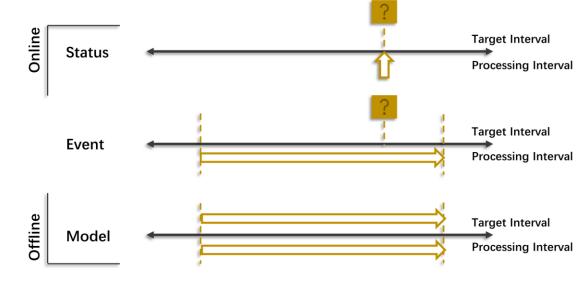
- Visual data
- Audio data
- Text data
- Sensor data
- Other data



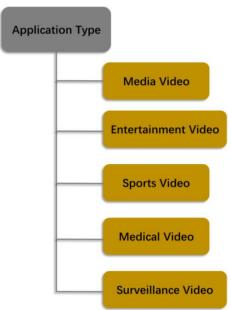
Video Data

□ Input State

Online || Offline



□ Application Type

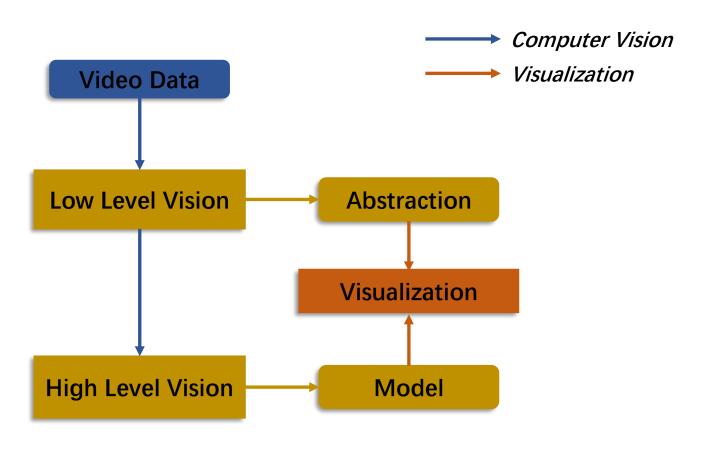


Video Analytics

□ Low Level VisionOptical Flow EstimationImage SegmentationFeature Extraction

•••••

☐ High Level Vision
Detection, Recognition, Tracking ……

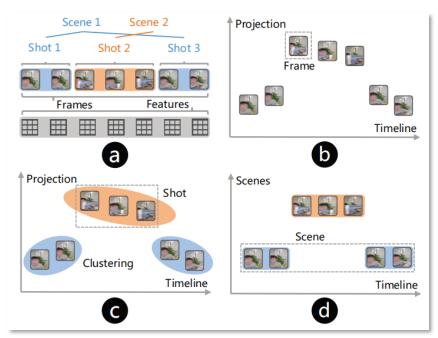


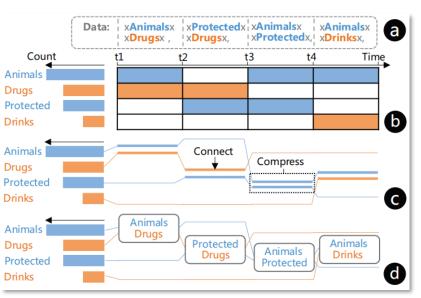
Outline

- 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





Frame View

(3)

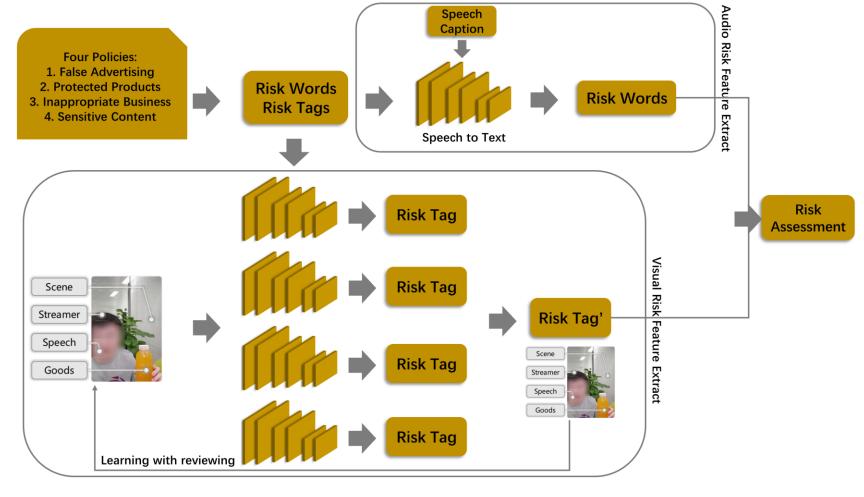
(1) Video Frame

(2) Audio Content

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

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.

Entertainment Video Vis I

Purpose

Explore, understand, and search movie content through the angle of emotion.

□ Target User

Audience and Editor

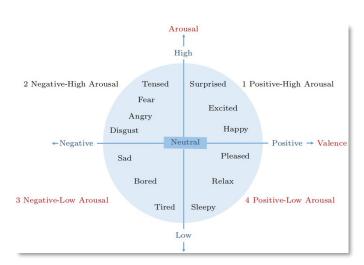
Data

Users' assessments of movies - Subjective

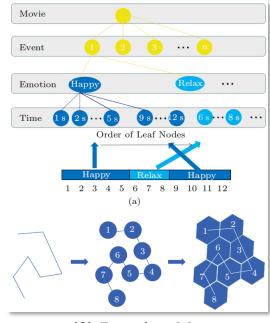
Characters' facial expression based on deep learning - Objective

Solution

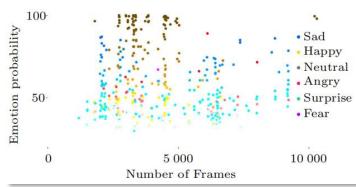
Visualization



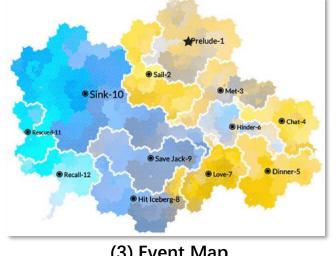
(1) Emotional Model



(2) Emotion Map



(4) Characters Emotion



(3) Event Map

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.

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

Emotional Factors	A Distribution 19 June 1	.124	E-factor Graphia	(B) (C) (C) (C) (C) (C) (C) (C) (C) (C) (C			(B) (12) (12) (12) (12) (12) (12) (12) (12
Emotional Factors Factor	nenees % Distribution % hard	.124	E-factor E-ropins Wordy.	B at Exercise	Total Composition of the Composi	auth)	- ther
Emotional Factors Factor 1 Effective Anousel_Culmenum Inskyrific Anousel_Mean Very sign	nenees % Distribution % hard	.124	E-factor E-spira	B E-sorpet	E-type E-simi	arity	- ther
Emotional Factors Factor 1 Effective Annual Culmenum Implifit Annual Mean Very sign	nenees % Distribution % hard	.124	E-factor E-spira World Sens	al E-script (0	- ther
Emotional Factors Factor 1 Effective Annual Culmenum Implifit Annual Mean Very sign	nenees % Distribution % hard	.124	E-factor E-spira World Sens	al E-script (0	0
Anousel_Culmenov Insignific Arousel_Mean Very sign	hverst grifficant and	.124	E-factor E-spira World Sens			0	
Arousal, Mean Very sign	grificant ant	.006	Sem/_	0 020020 01			
	ant	.020	Sem/_	0 020020 01			
Arousal_Volatility significar		. (R) ou				THE PARTY NAMED IN COLUMN TWO IS NOT THE PARTY N
Diversity Insignfic	cant	.120	Ohrles	00 000000000000000000000000000000000000		000	
Final_Arousal Very sign	grificant	.002	4.			° •	
Final_Valence Significa	ant	.020	-0.05	0.00 0.05 0.10			
Happy_Ratio Very Sign	gnficant	.001		Arousal_Mean	of speeds 4 SPS of speakers - 9		
Valence_Coherence Near Sig	gnificent	.051	Timeline	©			- 71
Valence_Mean Insignific	icant	.431	1-		volence /	arousal	Speaker
Valence_Volatility Insignific	icant	.207	- Halles As Anthonia	white the second	A STATE OF THE STA	 	sherrie
			- 1 sho	400 600	eós .	1,000	Speaker origin D88
Nonemotional Factor	rs		E-Spiral	E-script	E-type		Speech level
Factor 1 Effective	veness % Distribution %	Value 16		THE STEP IN THE STEP IN THE STEP	helitate and his	44	Final Speaker rank
Pauses Insignific	lcant	.271		OF LEGISLATION CONTROL	is historial the		2
Vocabulary Near sig	gnificant	.089		100 100 100 100 100 100 100 100 100 100	Diversity Emo-fin		Year 2018

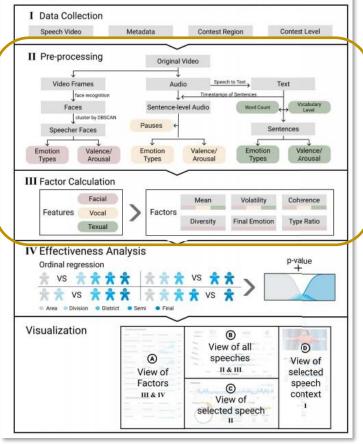
(1)

Factor	Modality	Type(p-value)	Type(p-value)			
	Facial	Arousal(0.006*)	Valence(0.431)			
Average	Textual	Arousal(0.215)	Valence(0.088)			
	Vocal	Arousal(0.016*)	Valence(0.017*)			
	Facial	Arousal(0.020*)	Valence(0.006*)			
Volatility	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)			
		Happy(0.001*)	Sad(0.0736)			
Datia	Facial	Fear(0.582)	Angry(0.292)			
Ratio	raciai	Surprise(0.115)	Disgust(0.306)			
		Neutral(0.488)	-			
Pauses	Vocal	Pauses(0.271)	-			
Vocabulary	Textual	Vocabulary(0.089)	-			

(2) Factors List

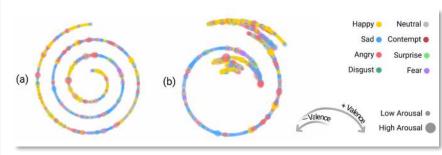
Entertainment Video Vis II

- Solution
- Visualization



Module	Description	
E-factor	To evaluate hypotheses of interest about speech factors using the cumulative data of all speeches.	Ī
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



DOES SHE LOOK A HUNDRED Y E S BUT

IF YOU MET GRAMMY SHE'S SWEET PETITE IN

HER SMILE LIGHTS UP A ROOM AND ANYONE

IN IT AS WE WERE CELEBRATING HER

MILESTONE BIRTHDAY AT THE NURSING HOME

I ASKED HER GRAMMY WHAT'S THE MEANIN

OF L I F E

Emotion Source

Arousal

Valence

Pause

(1) Overview

(3) E-spiral

(4) E-script

Entertainment Video Vis II

Cons

The definition of *Valence* and *Arousal* is not explained. The accuracy of model is not mentioned.

Sport Video Vis

Purpose

Analyze soccer videos to help analysts gain insights into player behavior and team tactics.

□ Target User

Team Sport Analysts

Data

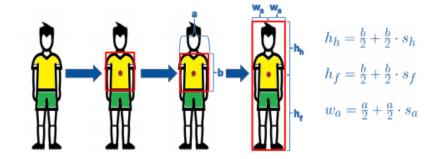
Soccer Match Video

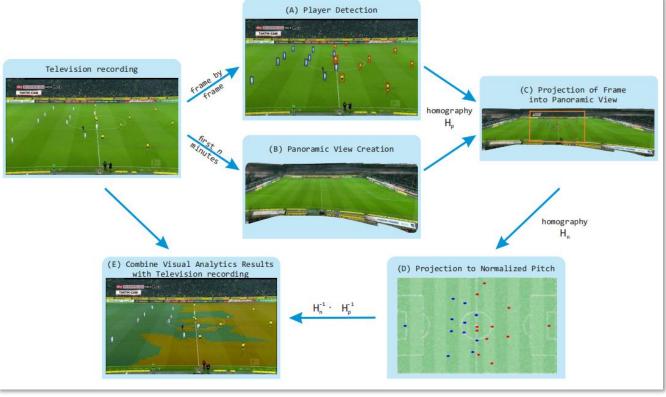


(1) Visual Analysis

Sport Video Vis

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





(1) Overview

Sport Video Vis

Cons

The tasks of domain experts are not rich enough. Event-based analysis is relatively simple. Color design conflicts.

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



Surveillance Video Vis

Purpose

Analysis of cheating behavior in online exams.

□ Target User

Teacher

Data

Mock Online Exam

Pitch

Cheating Types: Local Environment, Computer

Webcam Video Data --- Abnormal Head Movement

Mouse Movement (JavaScript Plugin) → Abnormal Mouse Movement

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

142121

(3)

Z

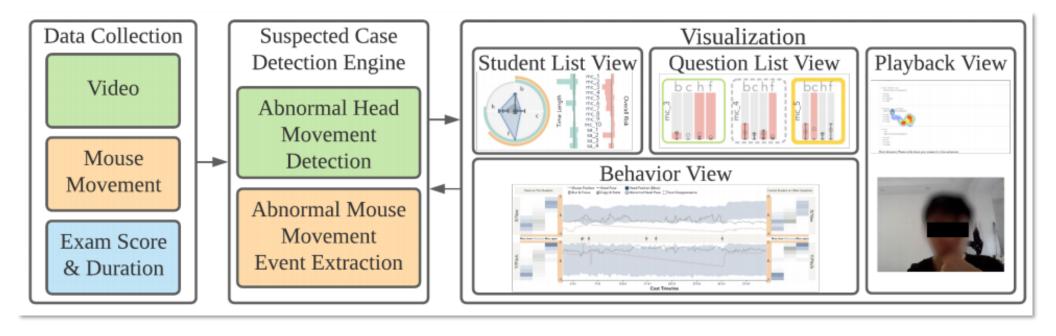
(1) Head Pose

(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

Related Papers

○ ○ ○ ○ ○ ○ ●

Surveillance Video

























-	A CONTROL OF THE PROPERTY OF T		7-00-	Management of the party of the	Constitution of the Consti							THE CONTRACTOR OF CAPACITY
	M	ledia		Enterta	ainment		Sp	ort	Medi cal		nce	
	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.
Info Meta	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
			(0)						[₹] (©			2 è
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		\triangleright					∇		\triangleleft			

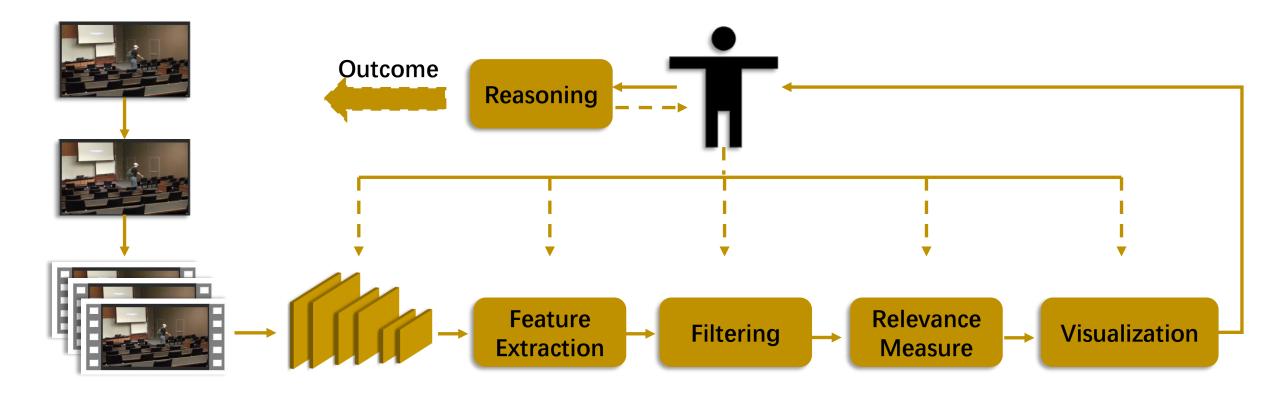
Outline

- 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.

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.