



IMVC 2024



Person tracker for Media and Entertainment Videos

Zvi Figov, Moti Kadosh, Yonit Hoffman PhD



Walking Next to People

PUBLIC Created 2 months ago by Zvi Figov

Embed Copy video ID Report Download Open in editor



main1 Appears in 94.38% of video



Play previous Play next

Observed people

111 observed people (preview)



Observed person 022
Appears in 4.19% of video
Short sleeves - Long pants



Play previous Play next

8 topics

Technology / Information systems / Videos on the web

Videos Music videos Video production

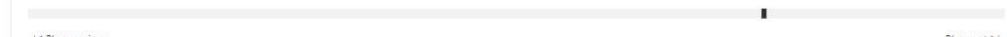
+4 more topics



Play previous Play next

1 audio effect (preview)

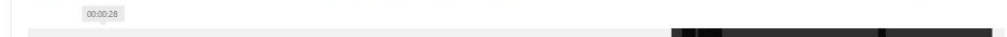
Laughter



Play previous Play next

23 keywords

daws video watch click walking people making social experiment camera totally kidding alright dude cool thought guys watching good -5

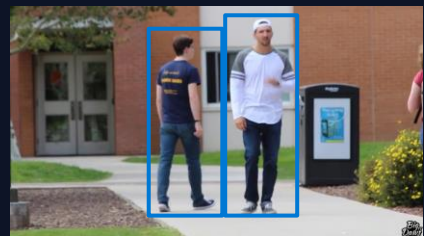


Play previous Play next

41 labels

Person Tracking Flow

For every frame



Inference models

ReId embedding



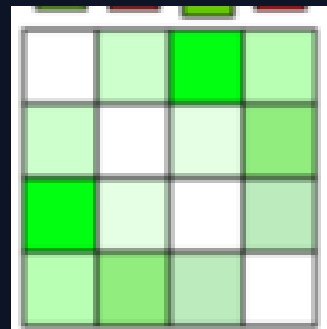
Track ID 22



Track ID 25



Association Matrix



Track ID 22



Track ID 25



Detection

Feature
Extraction

Affinity stage

Association

Track Management

Challenges specific to Video Indexer Tracker



6 sec clip -> 3 shots!

Editing effects



**Surveillance
vs M&E**



GPU CPU



**Compute
limitations**

Global shot detection

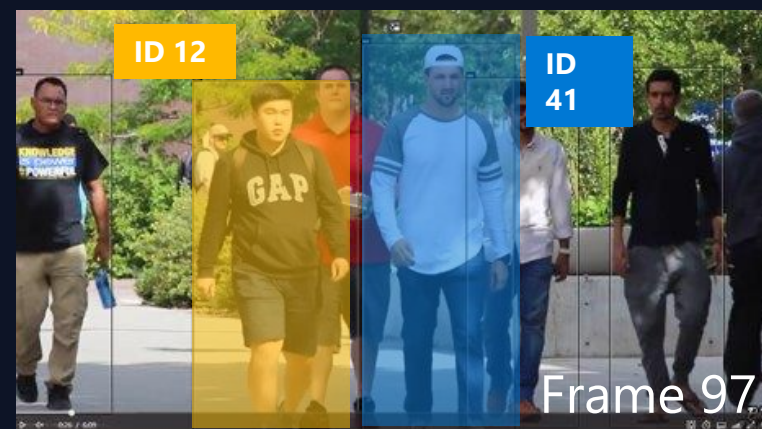
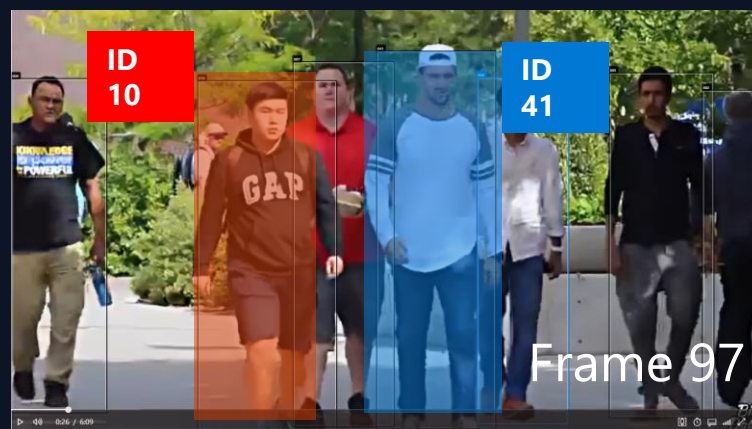
Tracker without shot detection

Tracker with shot detection



Frame 96

Frame 97

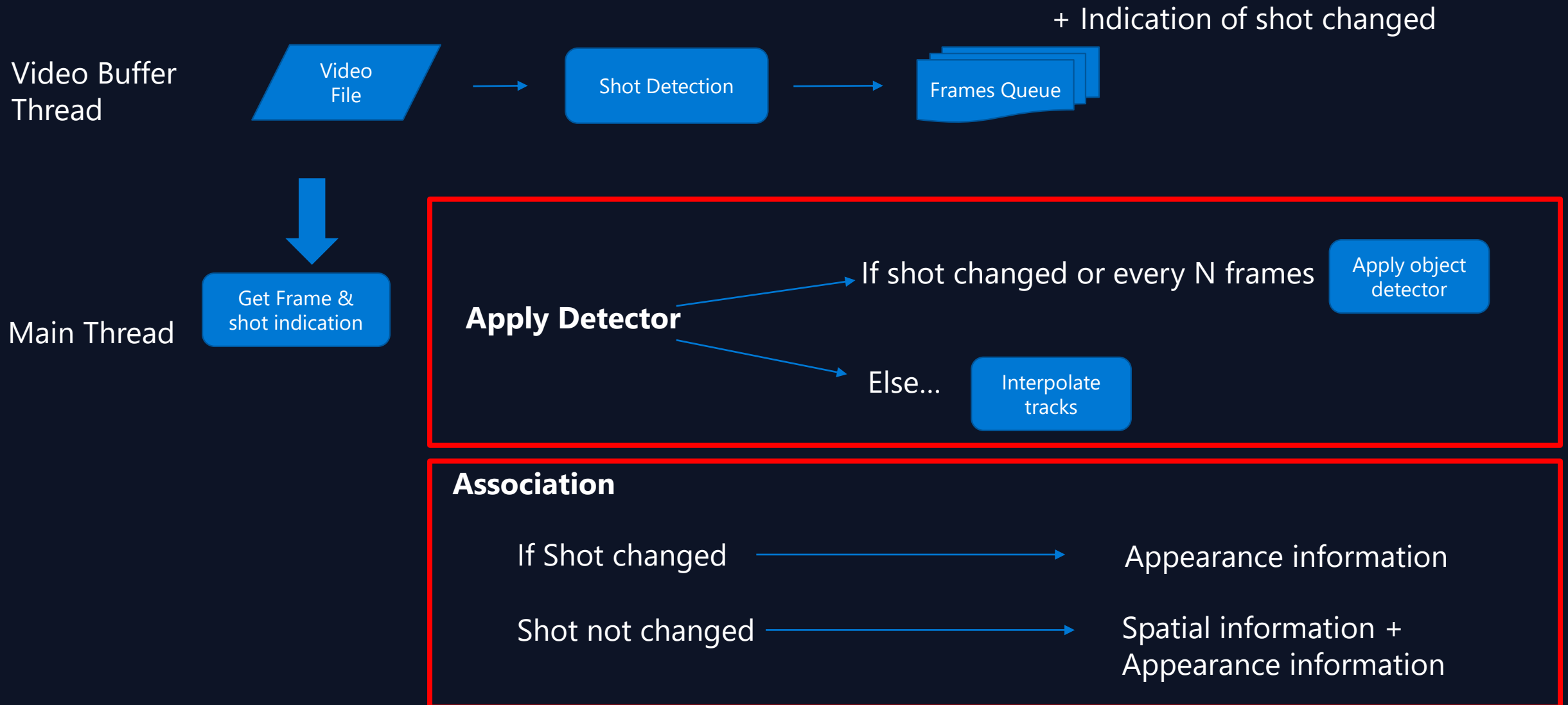


Frame 97

Frame 97

ID Switch

Pipeline for person tracking with the components

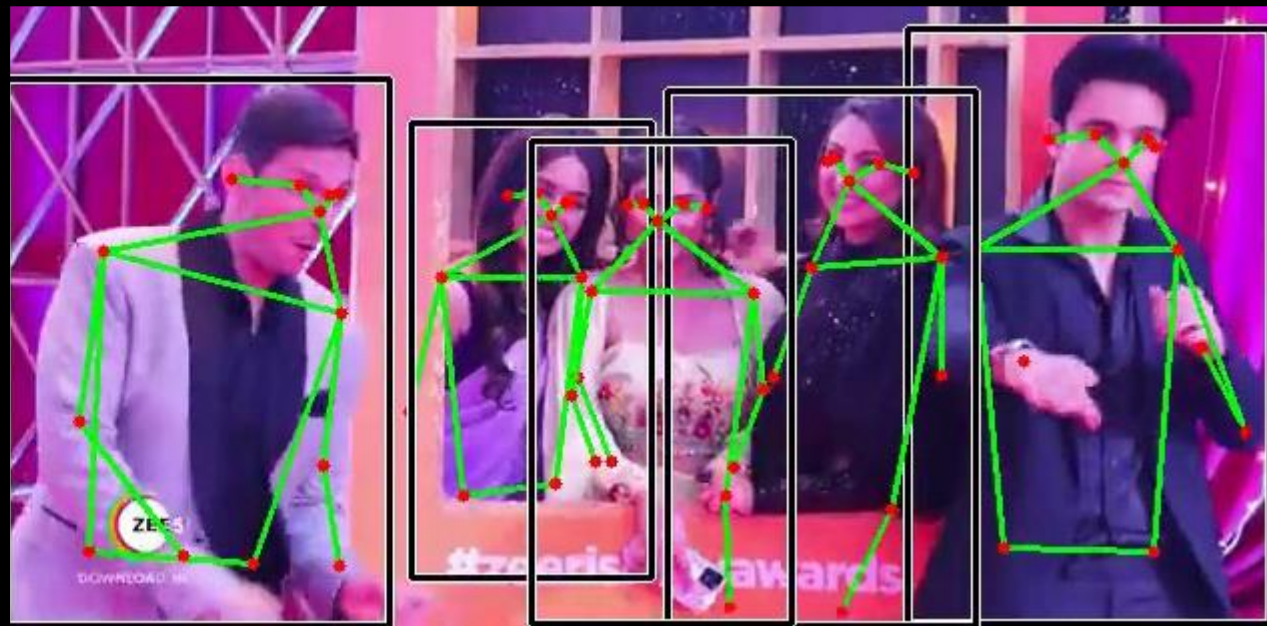


Additional Optimizations for Media

Detector – Full body

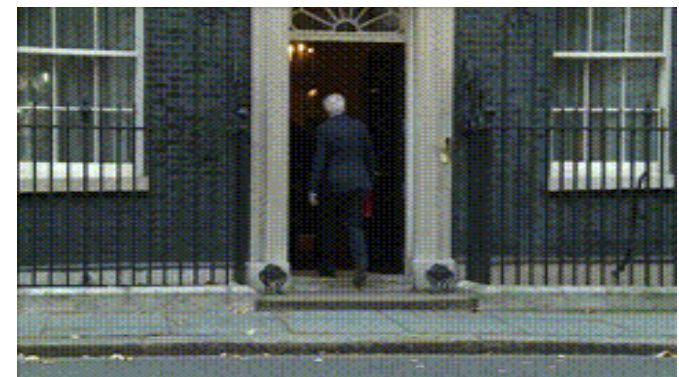
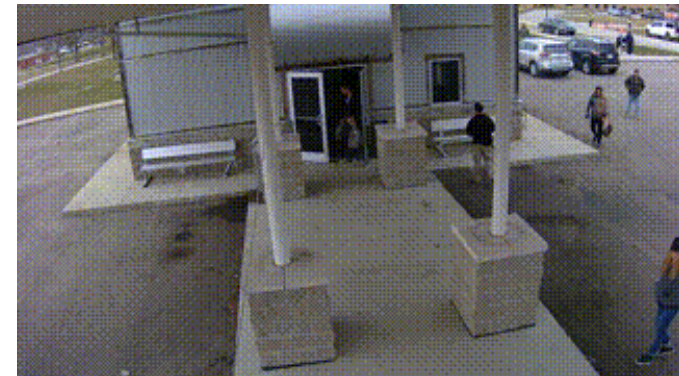


Detector – Human Pose Estimator



- Detector changed to pose (with key points)
- Filtering based on KP visibility
- Representative choice based on KP
- Parallel queue management for reading video

Dataset - 20 Diverse clips



Results

- Turnover duration = processing time / video length
- **ID switch** refers to the number of times a person's track receives different predicted IDs
- **Fragmentation** refers to the degree to which a person's track is broken up into smaller segments

	Detection only			Tracks quality		Processing time
	Recall	Precision	F1	Id switch	fragmentations	Turnover Duration
Baseline	0.51	0.83	0.63	877	2094	80%
Improved	0.44	0.92	0.60	430	1895	35%
overall diff	-7%	9%	-0.03	-50%	-10%	-56%

Summary

- We presented a novel pipeline for person tracking in videos that is specifically tailored for Media and Entertainment videos
- Improved results / reduced computation and cost
- Running in production at :

<https://www.videoindexer.ai/>