



Lightweight Client-driven Personalized Multimedia Framework for Next Generation Streaming Platforms

Ghulam Mujtaba

Advisors: Eun-Seok Ryu, Jaehyuk Choi

Gachon University, Seongnam, South Korea

Contents



1. Background
2. Overview and Universal Modules of Proposed Framework
3. Personalized Movie Trailer Generation Framework
4. Personalized Animated GIF Generation Framework
5. Personalized Keyshot Video Summarization Framework
6. Dissertation Conclusion

01



Background: Motivation and Introduction

Why do we need personalized media generation methods?

- Online video will make 82% of the global Internet traffic by the end of the year 2022
- 400 hours of video are uploaded to YouTube every minute
- Users have a “capture first, filter later” mentality
 - Later often means never in most cases



Getting information off the internet is like taking a drink from fire hydrant.
(Michell Kapoor)

01

Background: Motivation

What news should I read?

The screenshot shows the Yahoo! News homepage. At the top, there's a search bar and a navigation menu with links like Home, U.S., Business, World, Entertainment, Sports, Tech, Politics, Science, Health, Travel, and Most Popular. Below the menu, there are sections for Photos, Opinion, Local News, Odd News, Comics, Weather, Full Coverage, Video/Audio, Kevin Sites, and Site Index. A main headline reads "Severe Storms Leave 14 Dead in Midwest" with a photo of people in a damaged area. To the right, there's a video thumbnail for "WATCH VIDEO" and a link to "Winchester gun company shuts down". There's also a sidebar for "MORE STORIES" and "KEVIN SITES IN THE HOT ZONE". At the bottom, there's a "PHOTO HIGHLIGHT" section.

What game should I play?

The screenshot shows the App Store interface. At the top, there are navigation icons and a search bar. Below that, there are two large game banners: "TOP ELEVEN 2015" and "DUNGEON HUNTER V". Underneath the banners, there's a section titled "Best New Games" featuring icons for various games like Minions Paradise, Xenowerk, Inside Out Thought Bubbles, Bonza National Geographic, harmony 3, Dragon Jump, Dream Drop, Lines the Game, and Garfield Che Game of Feces.

Best New Games



What video should I watch?

The screenshot shows the Netflix homepage. At the top, there's a search bar and a navigation menu with links like Watch Instantly, Just for Kids, Personalize, and DVDs. Below the menu, there are sections for "Action & Adventure", "TV Dramas", and "Critically-acclaimed Foreign Movies". Each section displays a grid of movie and TV show covers.

What music should I listen?

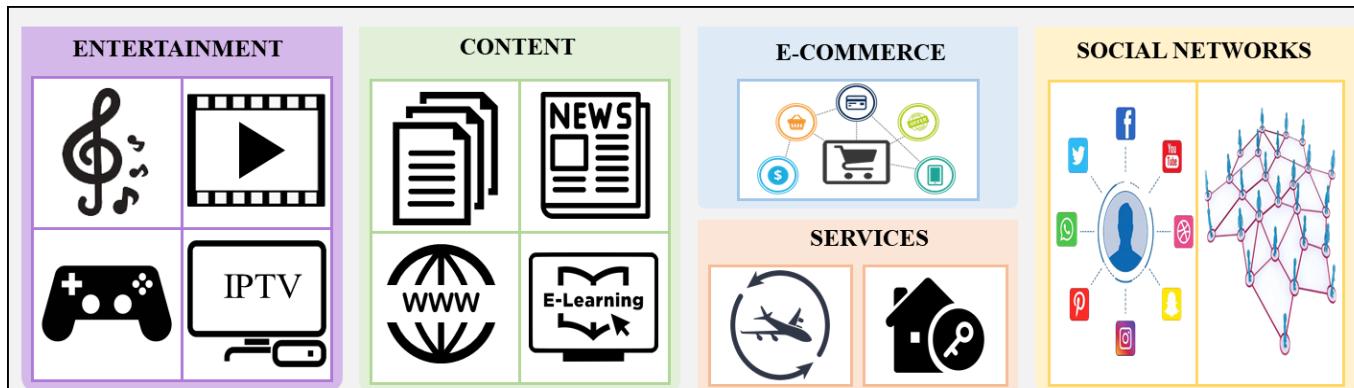
The screenshot shows a music streaming service interface. On the left, there's a sidebar with options like Browse, Discover, Radio, Follow, top Lists, Messages, Play Queue, Devices, and App Finder. The main area is titled "SORTED BY ARTIST" and shows a grid of artist and song covers. Some visible artists include BETTY WHO, BROODS, BRUNO MARS, LONDON, Heartbreak Dream, Broods, Bruno Mars, Because The Internet, DAN + SHAY, DAVID BISBAL, and DIEZ MIL MANERAS. There are also sections for "FOSTER THE PEOPLE", "MAGIC", "COLDPLAY", "EVERYDAY ROBOTS", "DAN + SHAY", "DAVID BISBAL", and "DIEZ MIL MANERAS".

Rise of Personalized Recommendation

- Provide content to **individuals** based on knowledge about their preferences and behavior
- The capability to **customize customer communication** based on preferences and behaviors at the time of interaction

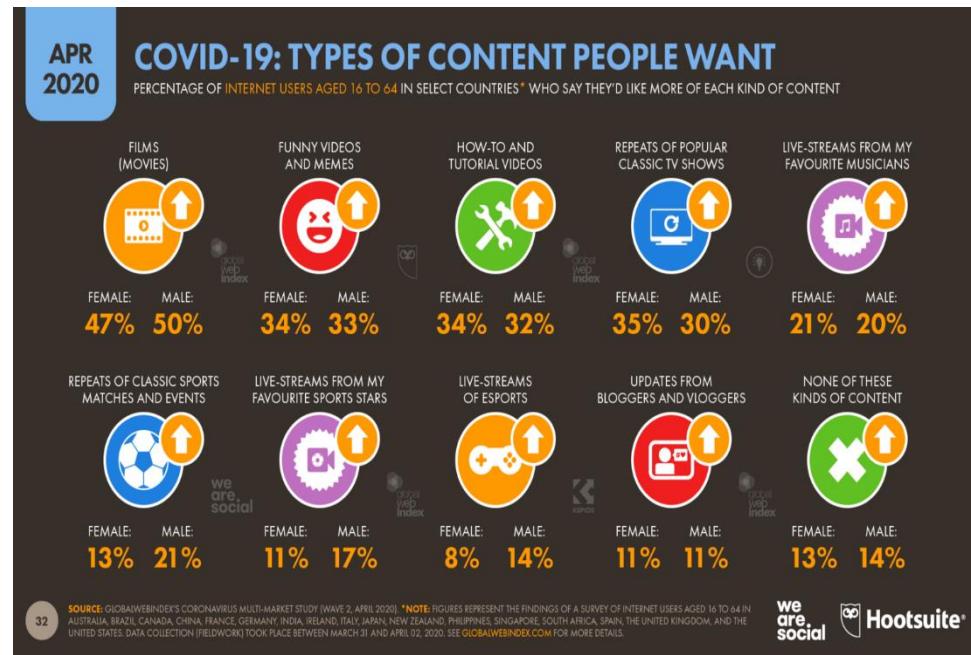
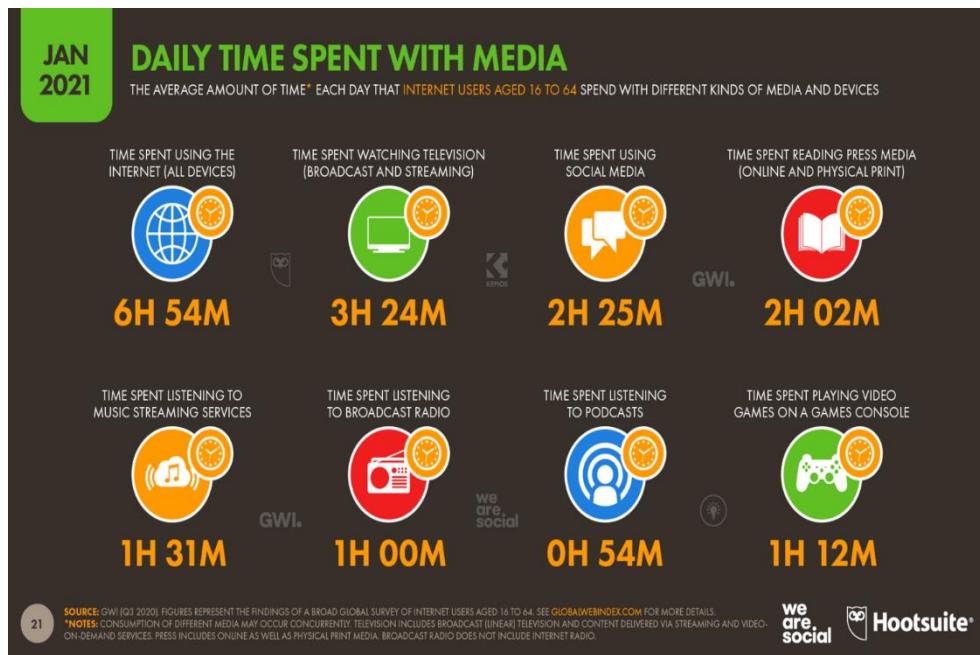
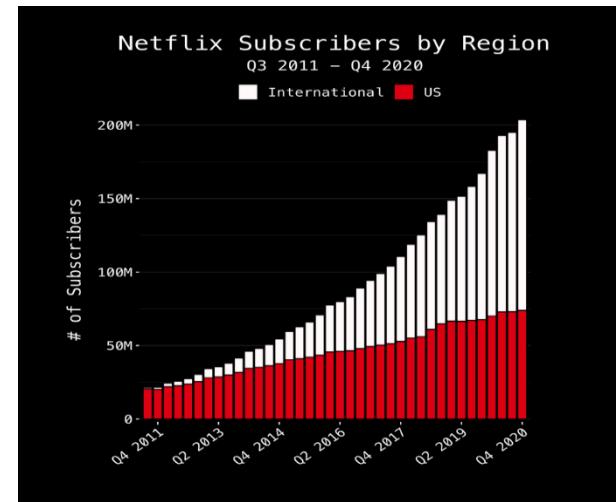


Applications



Rise of Streaming platforms

- Adults in the United States watch movies for nearly six hours each day (cisco)



Why do we need personalize media?

- Multimedia content generated one size fits all framework in streaming platforms
 - Trailer, animated GIFs, summary, etc.
- CTR important new videos
 - Economic cost if CTR reduce
- It is nearly impossible to satisfy all users on single generalize media (summary).
- Every user have interest and criteria even for the same video.



Mujtaba



Kim



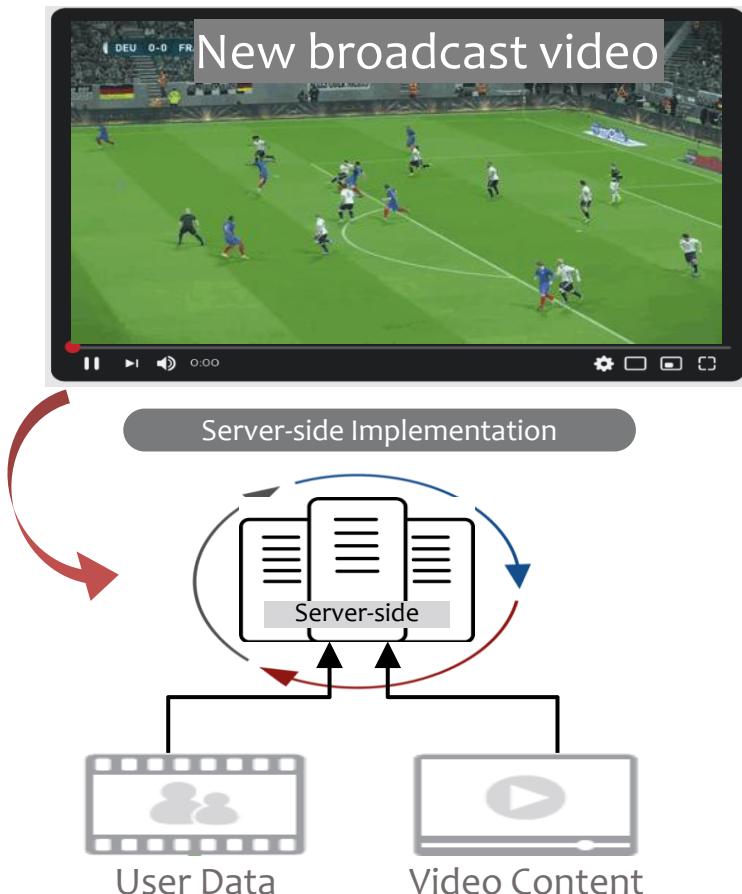
Why is it challenging?

- Human-centric task
 - Context dependent
 - Subject dependent
- What is multimedia depending on high-level semantics
 - Objects, actions
 - Motion, emotion
- Highly diverse inputs
 - Personalization should work in any setting
- Collecting ground truth is demanding
 - Prevents training large-scale models
 - Objectively comparing methods is difficult
- Provide real-time response



www.jolyon.co.uk

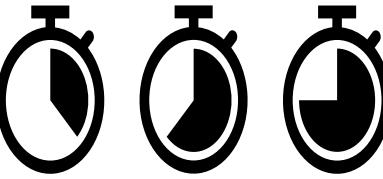
Traditional Solutions: [Server-based]



Challenges



User privacy



Optimal video length



Overwhelming user

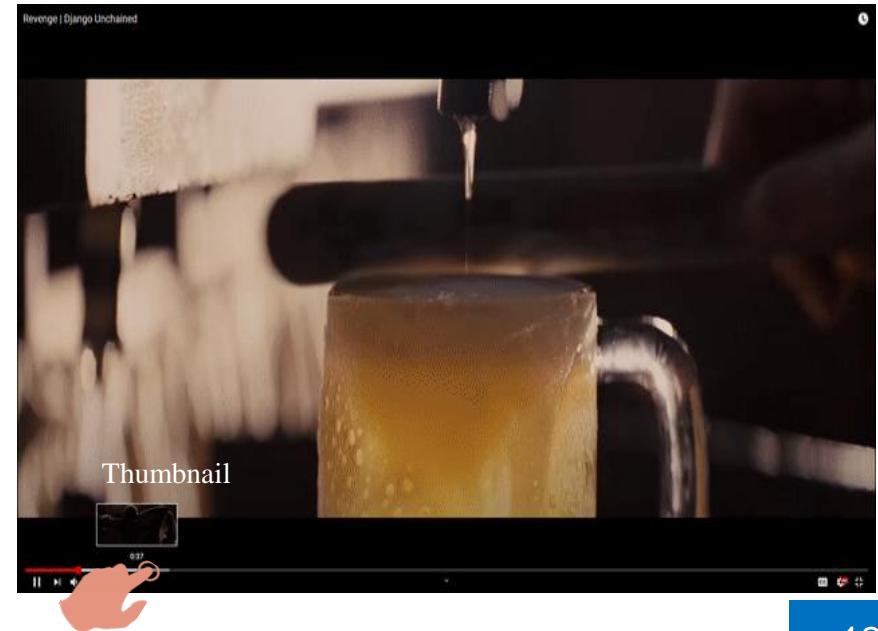
- Arduous to handle secondary/companion devices
- Challenging to get user emotions in real-time
- Oversee user behavior/ can control opinion
- Less scalable
- Wastes of computational resources during highly correlated frames processing

02

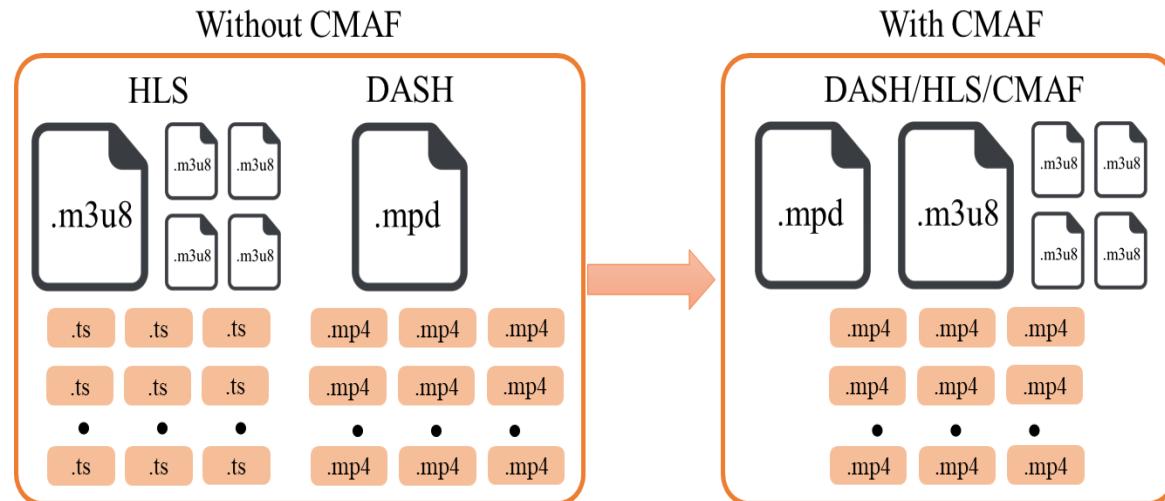
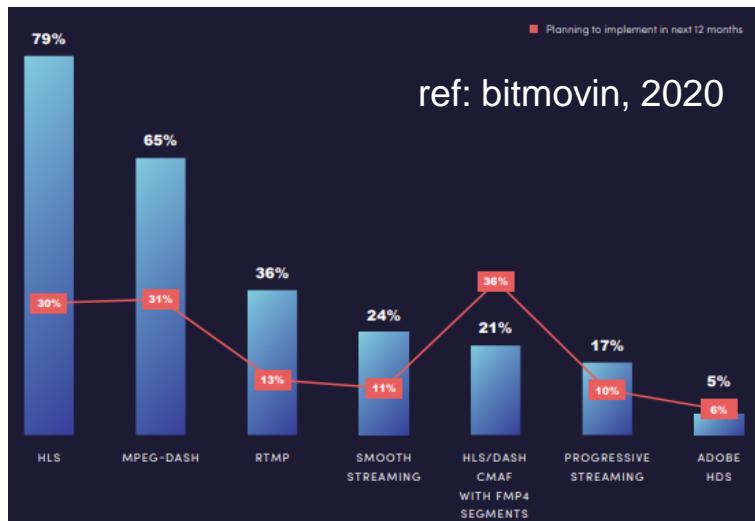


Overview and Universal Modules of Proposed Framework

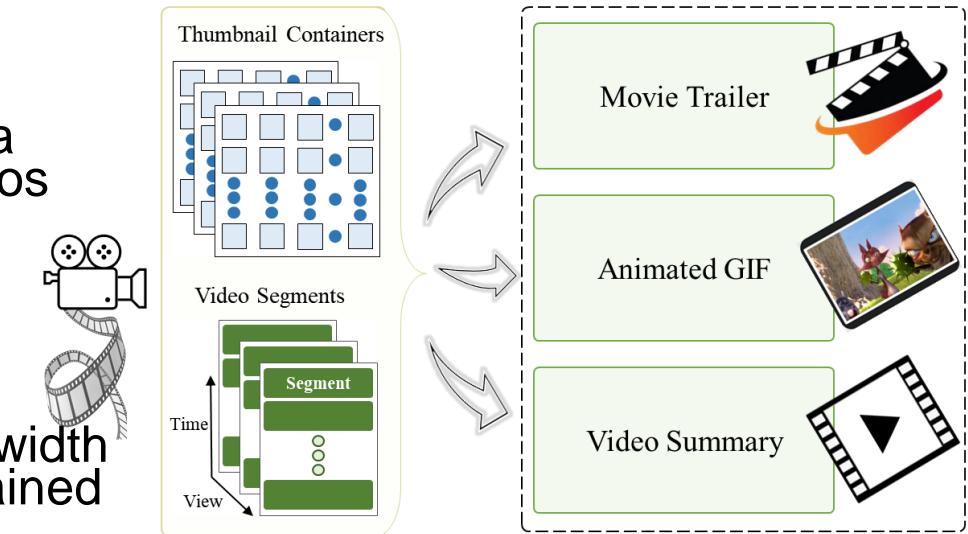
- Thumbnail
 - Used to video timeline manipulation
 - Seek and get and overview of full feature-length videos instantaneously
 - Easy to integrate in the browser
 - Lightweight (160x90) pixels
 - Easy to use, navigate, and generate
 - Widely used in most of the streaming platforms
- Thumbnail Containers
 - Combination of thumbnails, stored in nxm form. ($n=m$)
 - Here, each thumbnail represent 1 second, and thumbnail containers 25 seconds



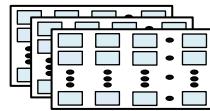
- The streaming platform services split the video into smaller video segments listed in the playlist file.
 - HTTP Live Streaming (HLS)
 - Dynamic Adaptive Streaming over HTTP (DASH)
 - Common Media Application Format (CMAF)
- HLS (.ts) segments can be decode and read/modify separately



- The first-ever client-driven thumbnail containers based personalize multimedia generation framework for long-form videos using resource-constrained devices
- Instead of processing entire video data (frames or video), thumbnails containers adopted
- Reduce computational complexity, bandwidth and storage demands resources-constrained client devices



Mujtaba: Scene/events Selection

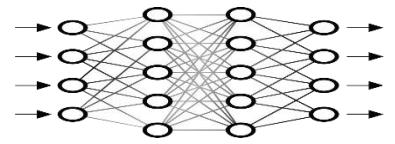


Input: Thumbnail Containers

User Preferences



Scene/actor network



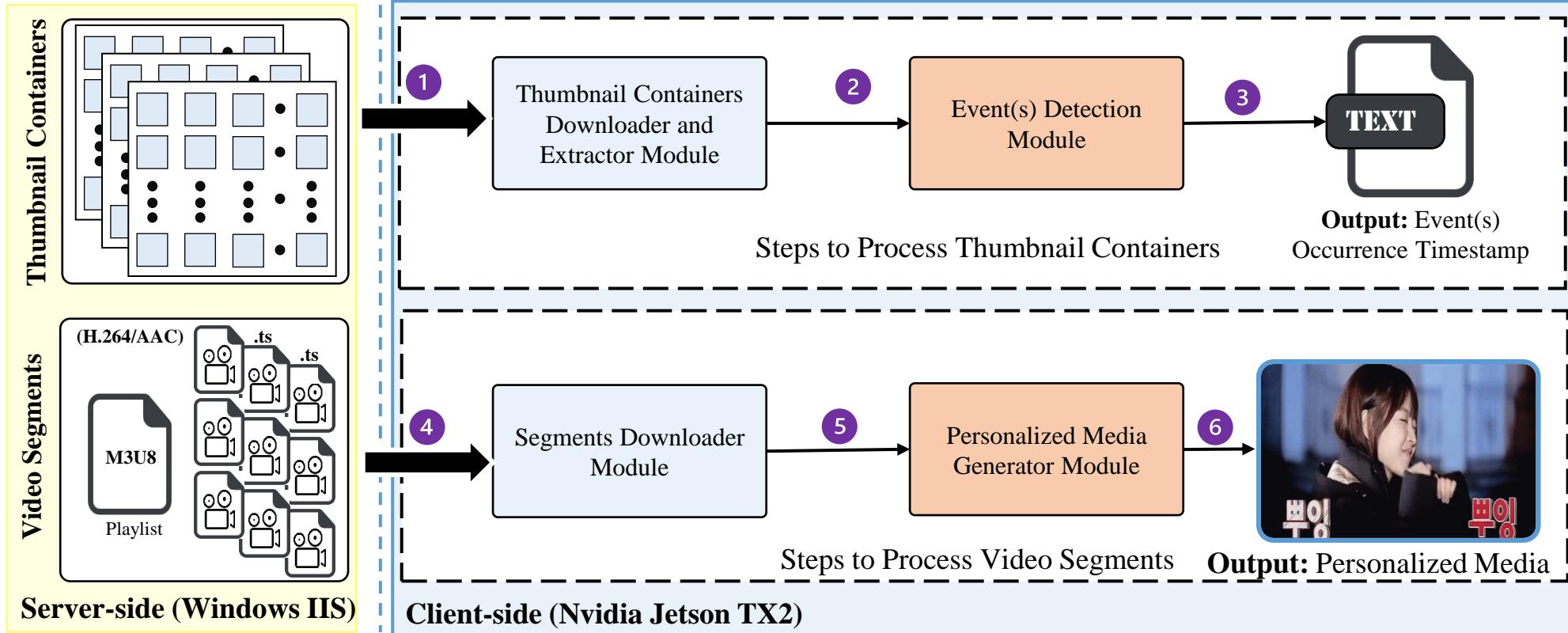
Results for Mujtaba



Kim: Scene/events Selection



Results for Kim



- General Modules
 - HTTP persistent connection
 - HLS Web-based video player
- Specific Modules
 - Event detection modules
 - Personalize media generation module

1 Select Video Title



2 Select Recommend Events

Personalized Media

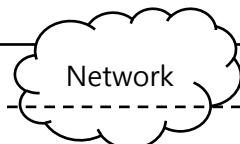


8 Segments Processing



HLS Server-side

3 Request Thumbnail Containers



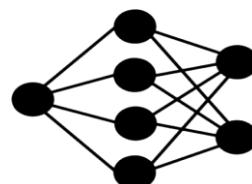
HLS Client-side

7 Sort List



6 Prepare List

5 Recognize Events



4 Thumbnail Extraction



Hardware Specifications

HLS Client

High Computational Resource (HCR)

- Ubuntu 18.04 LTS ()
- Quad-core 2.10 GHz Xeon
- GeForce RTX 2080 Ti
- 62 GB

Low Computational Resource (LCR)

- Nvidia Jetson TX2
- Nvidia Pascal 256 CUDA cores
- 8 GB

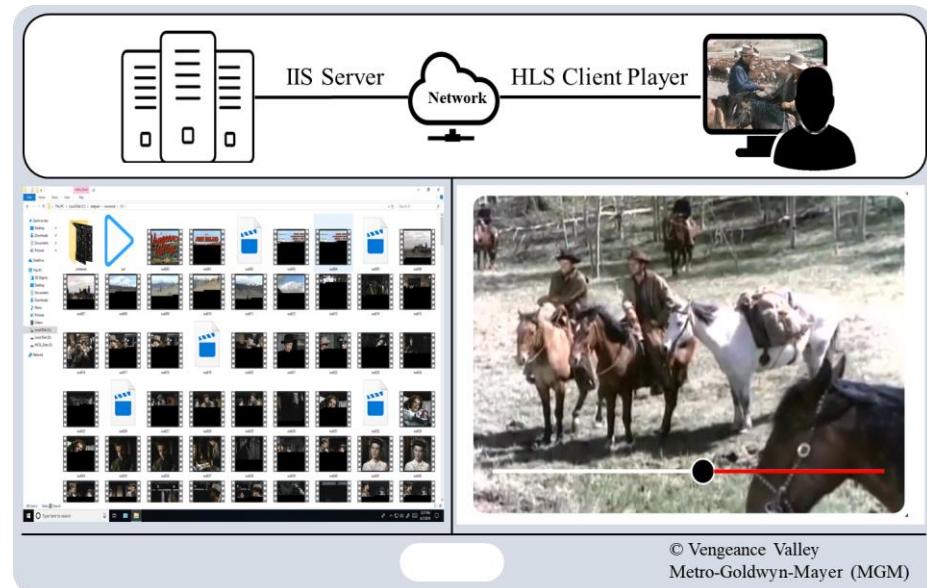
HLS Server

Windows 10

Internet Information Services (IIS)

Intel Core i7-8700K

32 GB



03



Personalized Movie Trailer Generation Framework

Why movie trailers are important?

- Makes the viewers more curious
- Provides opportunity to get to know a movie
- Helps you decide if a movie is worth watching
- Significance of Previews in Marketing
- Personal Importance of Previews



What makes it so difficult to generate a trailer?

- The trailer introduces the main characters
- The music gradually changes through the trailer, informing the audience about the direction of the story
- Elements such as lighting and editing work well with the music, also changing through the course of the trailer
- It should keep up expectations, and should appeal to the target audience

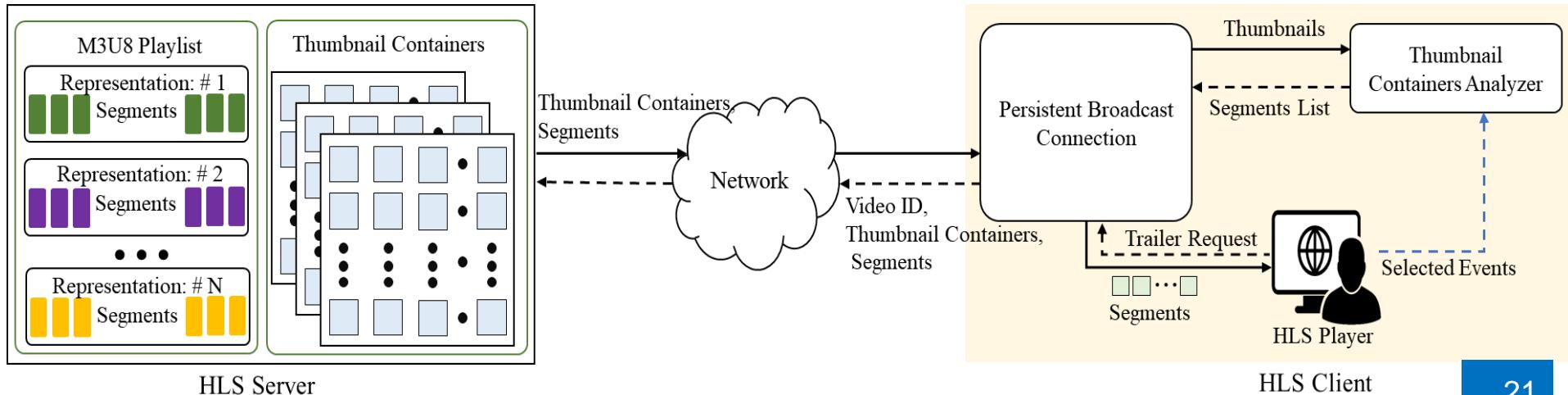
Motivation

- The first thing user wants to watch is trailer of the movie
- A handful of movie trailers are generated as the one-size-fits-all framework
- In some situations, a single trailer is generated for the corresponding movie
- Centralized server-side solutions are under consideration
 - Privacy breaches
 - Enormous demands for computational resources

Approach	Genre	Video category	Personalize	Data type			
				A	V	Tx	TC
Smeaton,et al [43]	Action	Movie	✗	✓	✓	✗	✗
Smith,et al [44]	Horror	Movie	✗	✓	✓	✗	✗
Kawai,et al [45]	Adventure	Documentary	✗	✗	✗	✓	✗
Proposed	Western, Sports	Movie, Documentary	✓	✗	✗	✗	✓

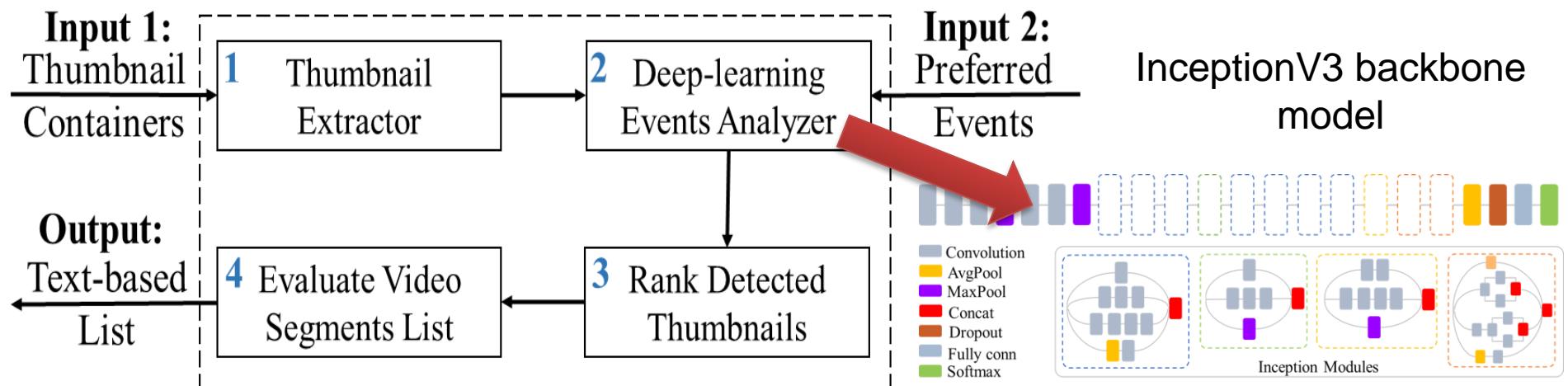
Proposed Method

- Facilitate personalized trailer generation process for full feature length videos
- Use of thumbnail, instead of video data (frames)
- Reduce computational and privacy bottlenecks
- Use computational resource of the end-user device
- Quantitative and qualitative assessments are conducted in 25 films and documentaries in the sports and western genres.



Process

- A movie with same genre, have similar events
- 6 different events selected
- Cricket, soccer and western genres videos analyzed



Experiments

S/N	FPS	# Frames	# Thumbnails	%
4	23	189983	7923	4.17
18	25	163972	6558	3.99
25	30	147448	4919	3.33

S/N	Baseline		Total (Baseline)	Proposed
	Frame extraction	Events recognition		
4	14.66	94.24	108.9	2.85
18	16.39	98.39	114.78	2.55
25	4.07	52.2	56.27	2.04

- 25 videos and trailers analyzed
- Proposed method **37.62** time computational efficient than baseline method

S/N	Movie Title	Genre	IMDB	Length	FPS	# Frames	# TC	# T	# Trailers	YouTube	Views
1	89 (2017)	Sport	7.8	1h31min	25	135,300	217	5412	1	UmUiHof01I4	127,335
2	Bobby (2016)	Sport	7.1	1h37min	25	1,420,204	225	5608	1	f1dJSOU-CUK	99,500
9	Bodyline (1984)	Sport	8.5	5h30min	25	73,820	119	2952	0	-	-
13	Bone Tomahawk (2015)	Western	7.1	2h12min	23	189,983	317	7923	1	QuGmtoQBPEM	498,587
14	Dakota (1945)	Western	6.1	1h22min	24	117,895	197	4917	1	CVeqwq-ZvVI	584
15	Django (1966)	Western	7.3	1h31min	24	131,749	220	5495	2	w8Ge2hmSTbo	1,420,015
16	Django Unchained (2012)	Western	8.4	2h45min	24	237,909	397	9922	3	eUdM9vrCbw	16,461,257
3	Goal! The Dream Begins (2005)	Sport	6.7	1h58min	23	169,932	284	7087	1	67LMX9-MHA	154,537
10	Iqbal (2005)	Sport	8.1	2h12min	25	189,945	304	7597	1	5Olc09k7KcE	121,231
4	Kenny (2017)	Sport	7.3	1h26min	25	129,649	208	5186	1	6mA6uA2-Rcw	219,456
11	Lagaan (2001)	Sport	8.2	3h44min	23	322,944	539	13469	1	oSIGQOYkFxs	363,132
17	Little Big Man (1970)	Western	7.6	2h19min	24	200,509	335	8362	1	7K4f5ZZe4-k	31,832
8	M.S. Dhoni The Untold Story (2016)	Sport	7.7	3h4min	24	265,933	444	11080	1	6L6XqWoS8tw	34,649,615
18	Oklahoma! (1955)	Western	7	2h25min	24	201,422	337	8401	1	V6uD9-aLCps	76,569
7	Pelé- Birth of a Legend (2016)	Sport	7.2	1h47min	23	147,545	257	6415	1	XBrfxHOXsDE	3,239,095
12	Playing Away (1987)	Sport	6.6	1h40min	24	146,237	244	6093	0	-	-
19	Shanghai Noon (2000)	Western	6.5	1h50min	23	158,648	265	6617	1	FqHg5fc_0_U	37,543
5	Take The Ball, Pass The Ball (2018)	Sport	8.2	1h49min	25	163,972	263	6558	1	Vfkls9Eo1Zl	421,157
6	The Game Of Their Lives (2005)	Sport	6.1	1h41min	24	145,699	244	6076	1	1H2fRn8PStw	128,803
20	The Indian Fighter (1955)	Western	6.4	1h28min	23	127,126	213	5302	1	hWP2Un2Dr5l	2,880
21	Vengeance Valley (1951)	Western	6.4	2h30min	24	148,404	248	6189	2	AlrWRttLTkg	848,745
22	The Rider (2017)	Western	7.4	1h44min	24	141,143	236	5886	1	2IV3LyS_M6M	92,368
23	The Tracker (2002)	Western	7.4	1h30min	23	132,192	221	5514	1	P0rK5Q-TX-k	4,150
24	The Train Robbers (1973)	Western	6.5	1h32min	24	158,733	265	6620	2	CUICu-zuAgM	5,977,606
25	True Grit (2010)	Western	7.6	1h50min	30	147,448	197	4919	1	j2srpV2RN-4	7,126

Results: Vengeance Valley

Similar Events in Official and Generated Trailer

Official Trailer

©Metro-Goldwyn-Mayer (MGM)



Running Time: 2min 34 sec
After: 0 to 11 and after: 21 sec



Running Time: 1 min 50 sec

Results: Oklahoma!

Similar Events in Official and Generated Trailer

Official Trailer

RKO Radio Pictures



Running Time: 3min 15 sec
After: 2min 20 sec



Running Time: 1 min 02 sec

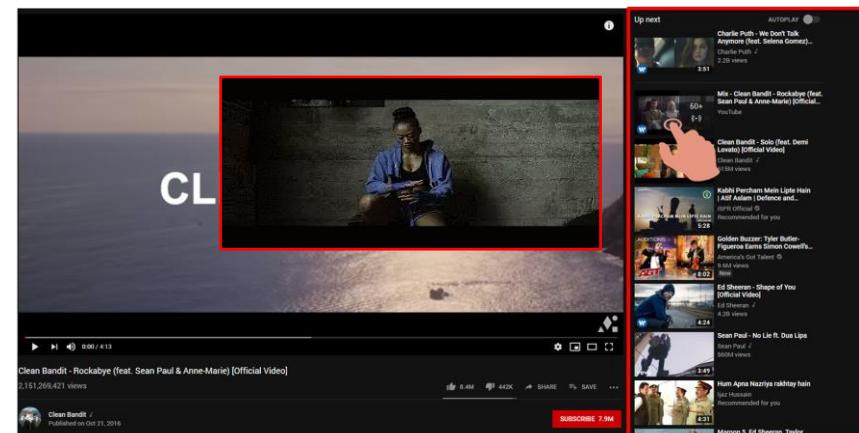
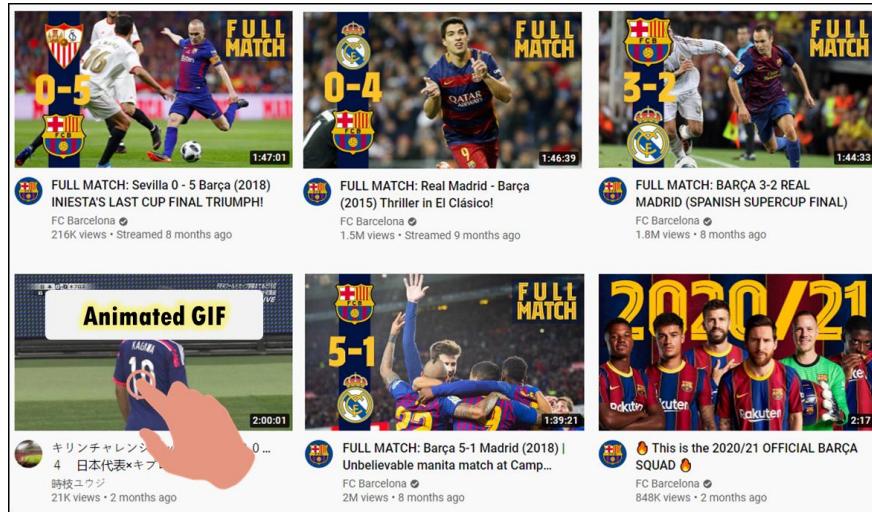
04



Personalized Animated GIF Generation Framework

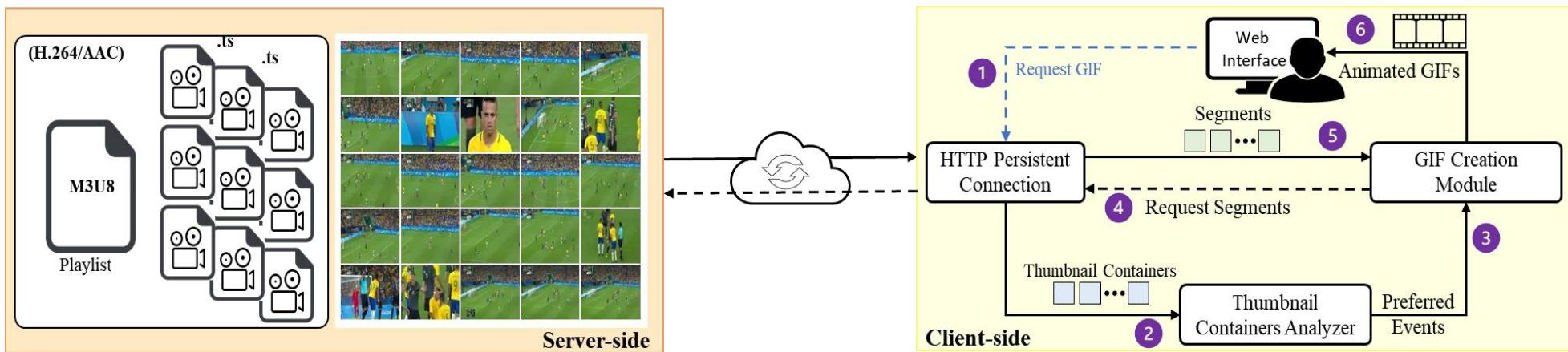
Motivation

- Animated images/video stick to your memory to make an impact -and to engage you!
- Usage of animated images
 - To tell a story
 - To show personality
 - To animate data
 - Quick Glimpse
 - Increase CTR
- Lightweight, no sound, loop
- Widely used in everywhere on the internet



Proposed Method

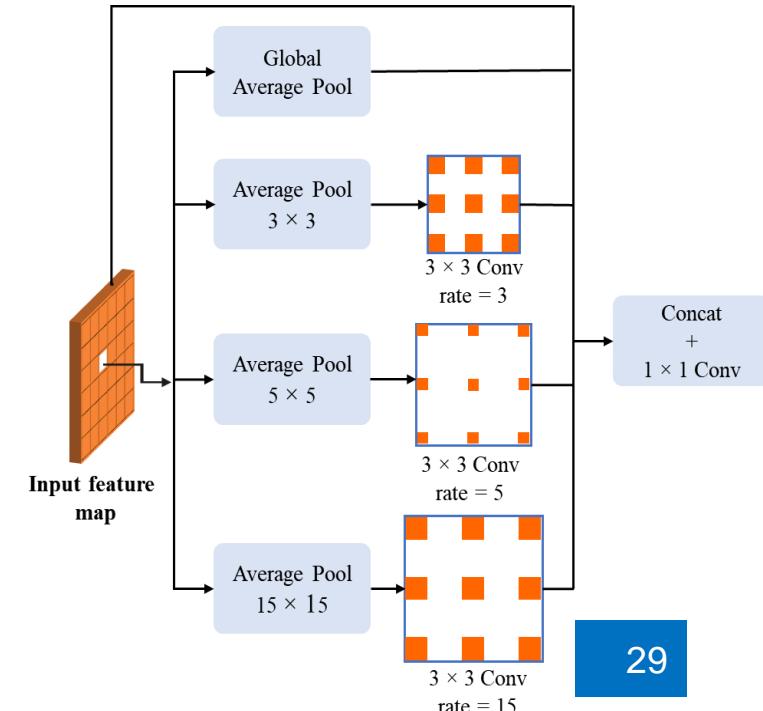
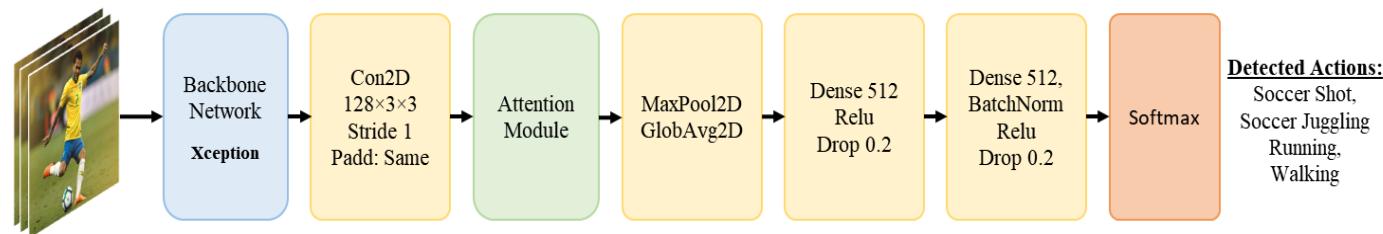
- A new lightweight client-driven method is proposed to create animated GIFs for sports videos (soccer)
- 2D CNN model is designed to identify personalized events according to user interests
- Extensive quantitative and qualitative analysis is performed using six sports videos. Quantitative results show that the proposed approach is **3.57** times more computationally efficient than the SoA GIF method.



Event Analyzer Module

- Designed 2D CNN network to analyze thumbnails
- Backbone Model
 - Xception
 - Vortex Pooling
 - SGDW optimizer
- TensorFlow (Keras)
- UCF-101 dataset

CNN Methods	Overall validation accuracy (%)
MobileNetV2 [69]	59.06%
MobileNetV3Small [70]	68.75%
MobileNetV3Large [70]	71.88%
DenseNet121 [71]	65.31%
InceptionV3 [57]	61.25%
Karpathy, Andrej, et al. 2014 [59]	65.40%
Shu, Yu, et al. 2018 [72]	76.07%
Mujtaba, et al. 2020 [10]	73.75%
Xception [65]	68.44%
Proposed	76.25%



Experiments

S/N	Title	Playtime	FPS	# Frames	# TC	# Thumbnails	Views	YouTube
1	Belgium vs Japan	1h 50m 50s	30	202,036	270	6734	1,141,707	ervkVzoFJ5w
2	Brazil vs Belgium	1h 50m 50s	30	199,506	267	6650	935,399	5OJfbYQtKtk
3	France vs Argentina	1h 50m 26s	25	165,653	266	6626	2,660,920	J41d0cHAfSM
4	France vs Croatia	1h 54m 1s	30	205,243	274	6841	1,367,451	7Fau-lwbuJc
5	Germany vs Mexico	1h 48m 56s	30	196,106	262	6536	1,111,419	3fYpcapas0k
6	Portugal vs Spain	1h 50m 25s	30	198,556	266	6625	1,792,000	Xhu5Bz1xDf0

HCR device

- **26.86** times faster than the HECATE
- **5.83** times faster than the AV-GIF
- **2.33** times faster Climax-GIF

LCR device

- 3.57 times faster than the Climax-GIF

Germany vs. Mexico

Video Size: 551 MB

Audio Size: 149 MB

ThumbCont: 22.2 MB

HECATE Yahoo Inc. 2016
Climax: MTAP 2021

S/N	HACATE [7]		AV-GIF [31]		Climax-GIF [31]		Proposed	
	HCR		LCR	HCR	LCR	HCR	LCR	HCR
1	85.34	16.69	38.71	6.83	10.08	2.19		
2	85.64	15.78	36.17	6.48	9.85	2.21		
3	81.86	15.56	35.40	6.42	9.32	2.25		
4	59.61	16.18	40.06	6.85	10.45	2.23		
5	57.81	15.35	37.96	6.57	13.96	2.09		
6	87.99	19.87	35.60	6.57	8.92	2.08		

Results

26.86 times

2.33 times

YouTube



HECATE



Climax



Proposed



Belgium v Japan



Brazil v Belgium



Germany v Mexico

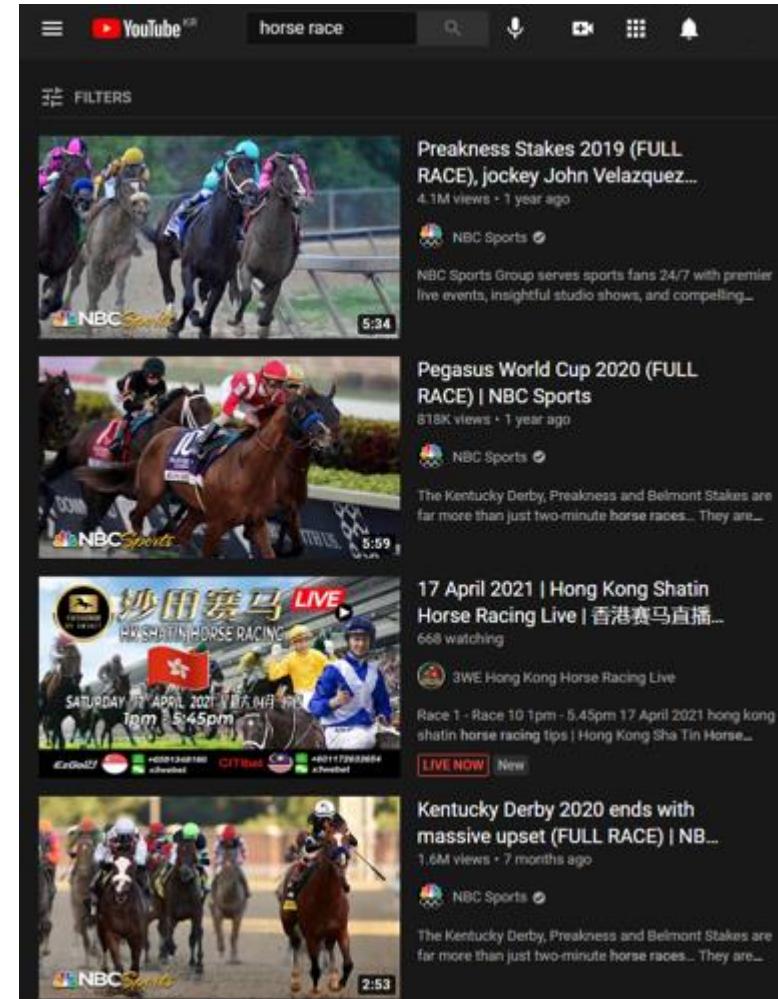
05



Personalized Keyshot Video Summarization Framework

Why do we need summarization?

- Large amounts of video
 - Need to search for relevant content quickly
 - Based on text search
 - Based on visual inspection
 - Time consuming
- Content quality degradation due to cheap acquisition
- Need better presentation
 - Through frame preview
 - Through automatic editing and summarization



What does it mean to summarize?

- Definition of “summarize” (text)
“Give a brief statement of the main points of (something)”
[Oxford dictionary]
- Similarly for video summaries
 - Brief
 - The summary should be non redundant or diverse
 - Cover the main points
 - Main: Frequently occurring content
 - Points: Interesting and visually informative frames or segments

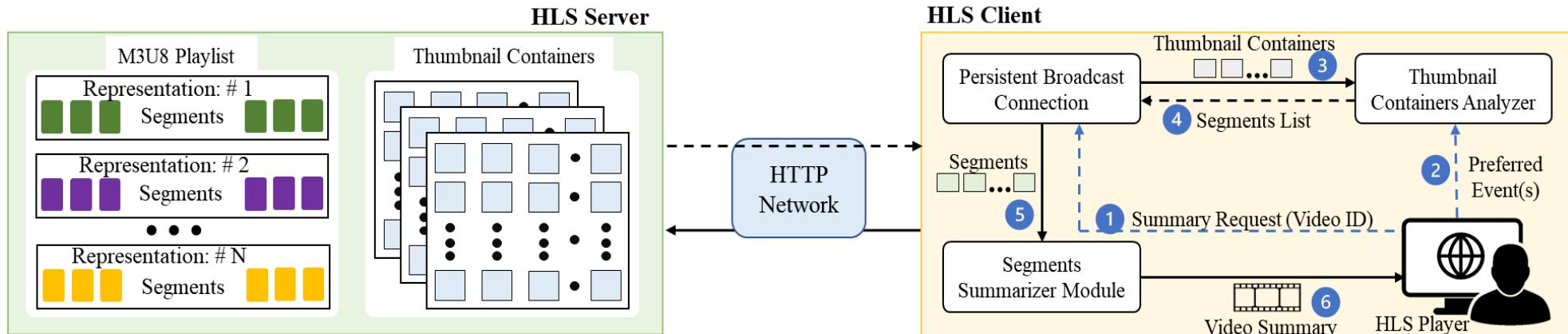
Summarization Types

- Keyframes (frames)
 - static storyboards, representative frames, or static image summaries
 - selects a small number of image sequences from the original video that presents approximate visual representation.
- Keyshot (short videos)
 - video skims, dynamic storyboards, or dynamic image
 - consist of typical continuous video segments of the full-length video that are shorter than the original video.



Proposed Method

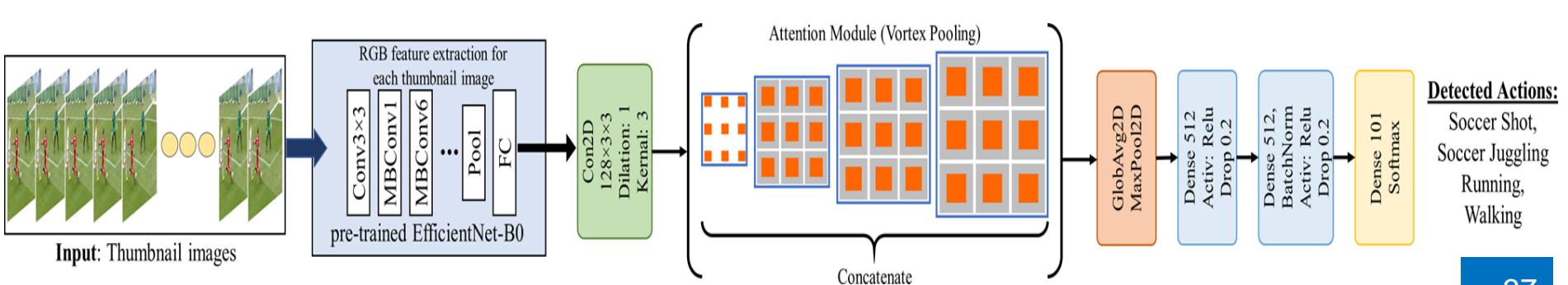
- A novel thumbnail-based client-driven framework is proposed to generate keyshot video summaries according to user preference.
- A lightweight 2D CNN model is designed that can identify personalized events from thumbnails.
- Quantitative and qualitative evaluations were conducted on eighteen movies and documentaries (approximately 32.9 h of duration).



Event Analyzer Module

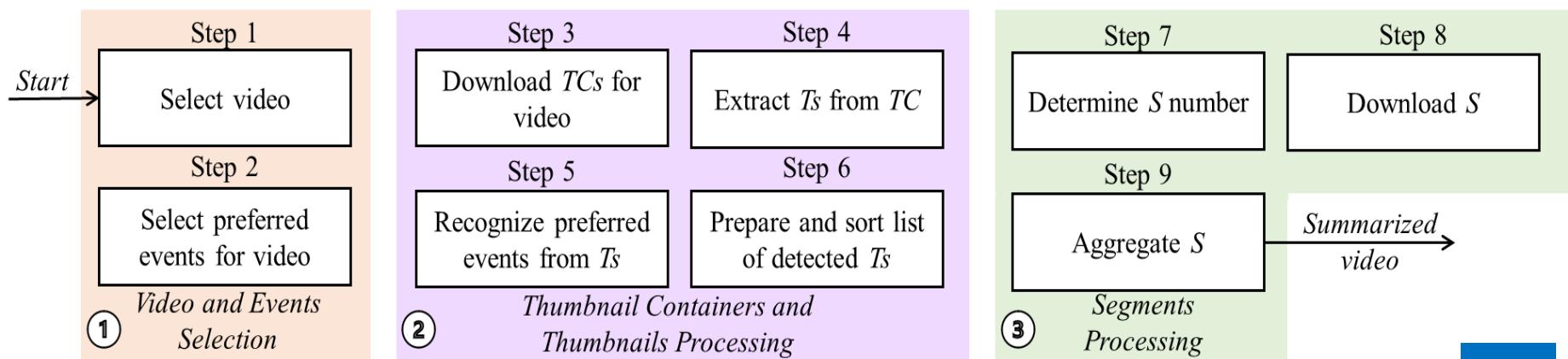
- Designed 2D CNN network to analyze thumbnails
- Backbone Model
 - EfficientNet-B0
 - Vortex Pooling
 - SGDW optimizer
- TensorFlow (Keras)
- UCF-101 dataset

Methods	UCF101
Karpathy, Andrej, et al. 2014. [59]	65.4%
Murthy, OV Ramana, et al. 2015 [58]	72.8%
Liu, An-An, et al. 2016 [78]	76.3%
Shu, Yu, et al. 2018 [70]	76.07%
Mujtaba, et al. 2020 [11]	73.75%
Mujtaba, et al. 2021 [79]	76.25%
Proposed	77.81%



Process

- A video with same genre, have similar events
- 18 videos analyzed
- 8 different events selected
- Sports, action and western genre's videos analyzed



Quantitative Experiments

S/N	Title	Genre	IMDB	Duration	FPS	# Frames	# TC	# Thumbnails
1	89 (2017)	Sport	7.8	1 h 31 min	25	135300	217	5412
2	Bobby (2016)	Sport	7.1	1 h 37 min	25	1420204	225	5608
3	Bruce Lee The Man And The Legend (1973)	Action	6.7	1 h 25 min	24	123658	207	5152
4	Django (1966)	Western	7.3	1 h 22 min	24	131749	220	5495
5	Django Unchained (2012)	Western	8.4	2 h 45 min	24	237909	397	9922
6	Goal! The Dream Begins (2005)	Sport	6.7	1 h 58 min	23	169932	284	7087
7	Little Big Man (1970)	Western	7.6	2 h 19 min	24	2005509	335	8362
8	M.S. Dhoni The Untold Story (2016)	Sport	7.7	3 h 4 min	24	265933	444	11080
9	Oklahoma! (1955)	Western	7	2 h 25 min	24	201422	337	8401
10	Shanghai Noon (2000)	Western	6.5	1 h 50 min	23	158648	265	6617
11	Snake In The Eagle's Shadow (1978)	Action	7.4	1 h 30 min	24	140473	235	5858
12	Take The Ball, Pass The Ball (2018)	Sport	8.2	1 h 49 min	25	163972	263	6558
13	The Indian Fighter (1955)	Western	6.4	1 h 28 min	23	127126	213	5302
14	The Legend Of Drunken Master (1994)	Action	7.6	1 h 42 min	24	147454	247	6150
15	The Rider (2017)	Western	7.4	1 h 44 min	24	148404	236	5886
16	The Train Robbers (1973)	Western	6.5	1 h 32 min	23	132192	221	5514
17	The Way Of The Dragon (1972)	Action	7.3	1 h 30 min	24	142712	239	5953
18	Vengeance Valley (1951)	Western	5.9	1 h 23 min	30	147448	197	4919

13.59 times faster than the HECATE

2.45 times faster than the DR-DSN [AAAI 2018](#)

2.42 times faster VASNet [ACC 2018](#)

2.38 times faster AC-SUM-GAN [IEEE Trans. CSVT 2020](#)

35 times faster FB-SUM

S/N	HECATE [6]	DR-DSN [8]	VASNet [9]	AC-SUM-GAN [10]	FB-SUM	Proposed LTC-SUM	
						HCR	
1	20.85	4.92	5.09	4.81	70.41	7.64	1.98
2	22.02	5.15	4.77	4.87	67.23	8.16	1.91
3	28.51	4.10	4.31	4.43	60.91	5.56	1.98
4	21.05	5.13	5.32	4.74	63.67	8.25	2.10
5	54.41	8.58	8.36	8.69	132.32	14.64	3.58
6	28.56	6.32	6.18	5.92	80.60	10.03	2.43
7	51.91	7.32	7.03	7.04	102.16	12.07	3.02
8	85.23	9.53	9.93	9.52	123.29	15.87	4.22
9	53.49	7.74	7.91	7.98	118.60	12.23	2.93
10	32.03	5.51	5.57	5.64	82.65	9.85	2.36
11	25.50	5.24	4.91	4.86	73.14	6.49	2.23
12	28.29	6.01	5.88	6.16	131.83	9.87	2.31
13	18.71	5.21	4.75	4.64	65.36	7.90	1.95
14	31.33	5.23	5.23	4.98	77.21	6.78	2.27
15	17.20	5.57	5.09	4.97	73.45	9.08	2.32
16	23.11	4.78	4.77	4.81	60.82	7.93	2.06
17	29.00	5.39	5.29	5.20	78.08	6.67	2.25
18	24.59	5.55	5.57	5.08	74.36	7.64	1.91

89 (2017) 1h 31 min (640 × 480 px)

Video Size: 612 MB | Extracted frames (1.1 GB)

ThumbCon size: 14 MB

Qualitative Evaluations

- 56 participates
- 9 geographic locations
- 12 different combinations of the events
- Entirely anonymous
- Similar events but higher ratings
 - Small summary duration and precise

Questions	Baseline	Proposed
Q1: Did the generated summary give related actions (events) according to your preferences?	7.14	7.59
Q2: Rate generated summary.	7.16	7.52
Q3: Is the length appropriate for the generated summary?	6.45	7.39
Q4: Compare to both generated summaries which one is good rate, please.	6.89	7.32
Q5: Correlations (similarities) of the generated summaries.	6.89	
Q6: Would you like to watch the movie after watching the generated summary?	7.09	7.14

Generated Summaries Results

2 × Speed

Original Video 2 h 45 min



Baseline
18 min 14 sec

Original Video 1 h 49 min



Baseline
22 min 40 sec

Original Video 1 h 30 min



Baseline
34 min 36 sec



Proposed
4 min 36 sec



Proposed
11 min 41 sec



Proposed
13 min 1 sec

06



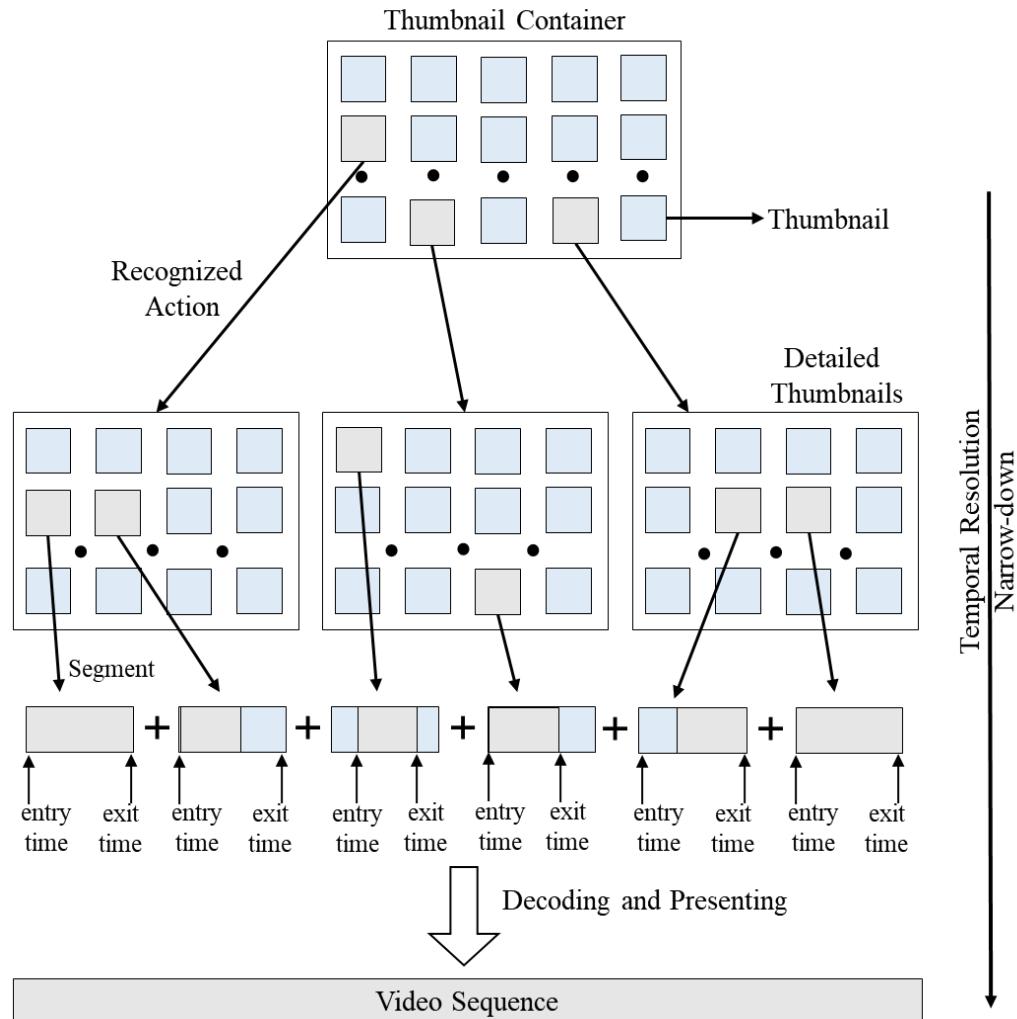
Dissertation Conclusion

Conclusion

- Proposed a lightweight client-driven personalized multimedia generation framework for streaming platforms.
- Long-form videos, reduce privacy and computational bottlenecks for resource-constrained end-user device
- By adopting thumbnails, reduce bandwidth and storage demands
- Three different client-driven techniques designed for streaming platforms to validate the proposed framework.
- The first proposed method is designed to facilitate and expedite personalized trailers generation in the film-making process.
- The second proposed method is designed to generate personalized animated GIFs for full-length sports videos.
- Finally, the third proposed method is designed to produce personalized keyshot-based video summaries for different video categories such as documentaries, movies, and sports matches.

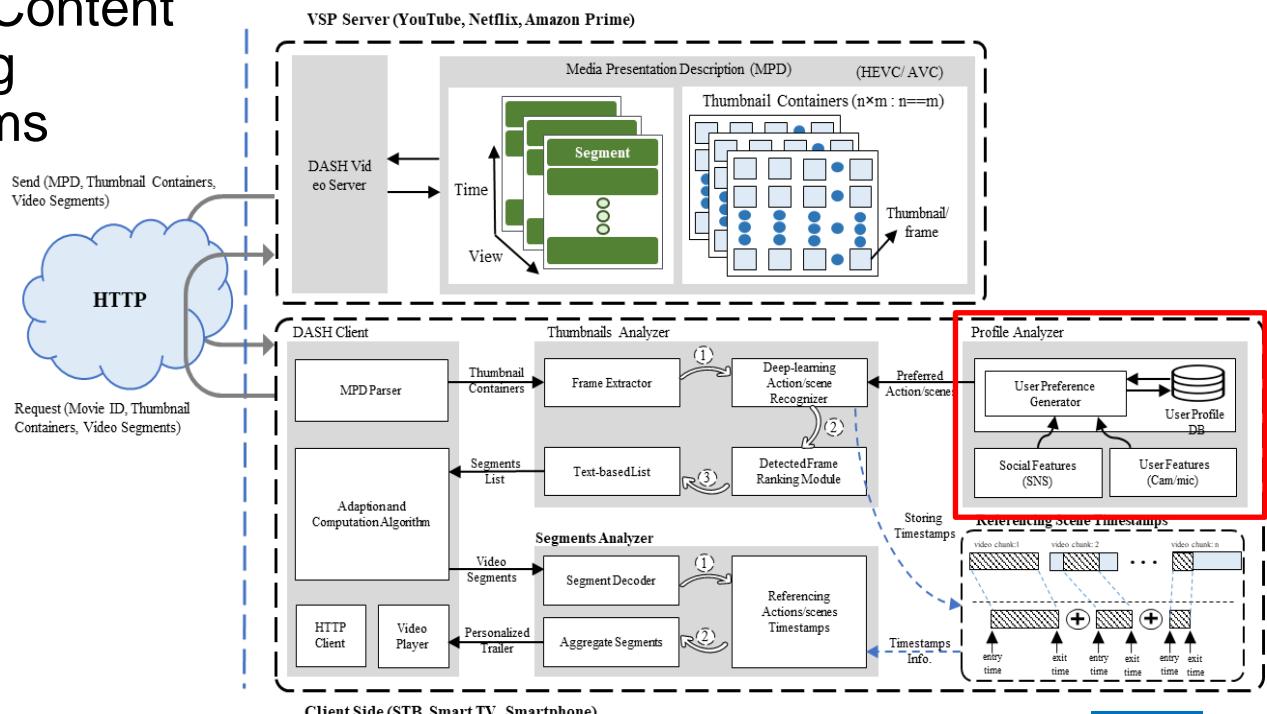
Limitations

- Unrelated Frames
- Maybe not efficient
 - for short-form videos
 - detect fast scene transition
- Need to investigate



Future Work

- Enhance computational efficiency by adopting hierarchical thumbnails
- Personalized Services for ATSC 3.0 using over-the-top (OTT)
- Personalized Multimedia Content Generation Methods using recommendation algorithms



THANK YOU