

Supplementary Material for ***TVSum: Summarizing Web Videos Using Titles***

Yale Song, Jordi Vallmitjana, Amanda Stent, Alejandro Jaimes
Yahoo Labs, New York

{yalesong, jvallmi, stent, ajaimes}@yahoo-inc.com

This supplementary material includes detailed information about our experiments presented in Section 5.2. Below gives a summary of the material; see captions in each Table and Figure for more detailed explanations.

- Figure 1 shows a tutorial page and the annotation user interface of our Amazon Mechanical Turk (AMT) study.
- Table 1 shows the original titles (provided in the SumMe dataset) and our substituted titles, along with YouTube unique video identifiers (some are missing because we could not find them).
- Figure 2 and Figure 3 show a list of 50 videos in our TVSum50 dataset.
- Figure 4 shows per-video F_1 scores on our TVSum50 dataset.
- Figure 4 shows detailed experimental results (show importance scores and visualizations of the learned co-archetypes) on our TVSum50 dataset.

Create a visual summary of a video

Our goal is to build an automatic video summarization system. One way to summarize a video is to combine chunks that contain the most important visual contents. For example, we can extract chunks whose visual content is highly relevant to the main topic and that displays the topic content in the most representative way (e.g., the best camera angle); the summary is then a chronological concatenation of the selected chunks. Your answers will help us gather data for that purpose.

Instructions:

1. Read the title of a video, try to guess its main topic
2. Watch the entire video
3. Evaluate the importance of each chunk
4. Answer additional questionnaire

How to score each chunk:

Please assign a numeric score to each chunk, on a scale from 1 (completely not important) to 5 (absolutely important).

Determining the importance of chunks is an ambiguous and subjective task. We want you to use your best judgement: think about how much relevant and representative it is of the topic content compared with other visually similar chunks.

Since not all parts of the video can be equally important, you will be allowed to assign a given score to only a limited number of chunks. For example, if you give parts of the video a score of 5 (most important), you can assign that score to only up to 5% of the video.

To that end, we set the target range of each score as:

- Score 5: 1% - 5%
- Score 4: 5% - 10%
- Score 3: 10% - 20%
- Score 2: 20% - 40%
- Score 1: 25% - 65%

For example, if a video has 100 chunks, you can only give a score of 5 to between 1 and 5 chunks (1%-5% of the video), a score of 4 to between 5 and 10 chunks (5%-10% of the video), etc.

Check out the [tutorial](#) for more details.

Hope you enjoy our task. Questions? Please contact us at mturk.guru.cjswo@gmail.com

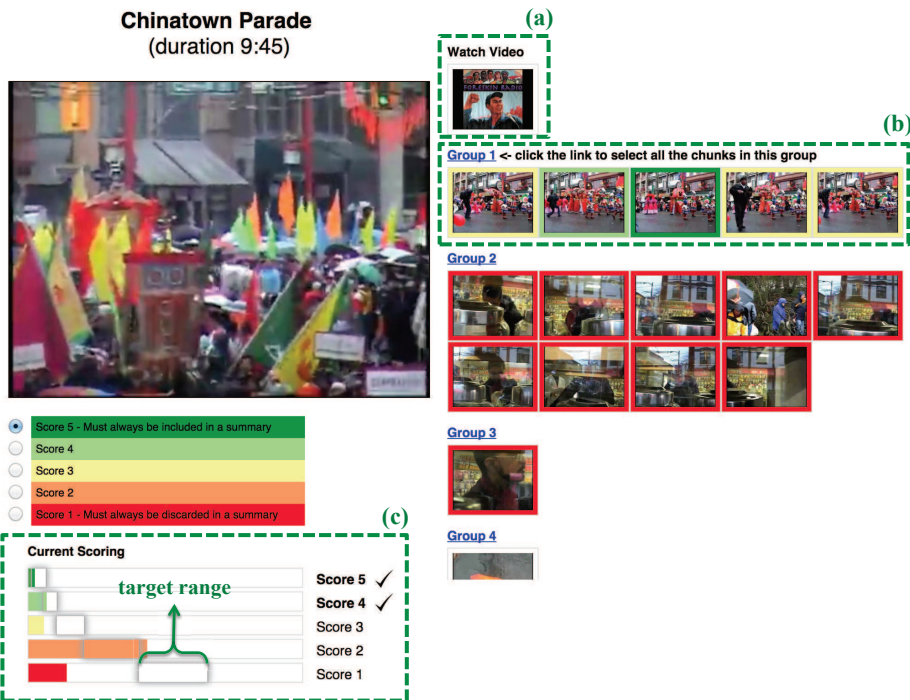

















Figure 1. **Amazon Mechanical Turk study user interface.** A tutorial page (top) and the annotation user interface (bottom). (a) By clicking the image below “Watch Video,” users can replay the video at any time during the task. (b) A group of visually similar shots; we clustered shots using the k -means algorithm. Users can assign a score to each shot in a group (by clicking a thumbnail image), or all shots in a group as a whole (by clicking the group header). The groups and shots are presented in a random order to avoid the chronological bias. (c) A monitoring gadget showing the current progress towards the target range of each score.






Table 1. **Substituted titles on the SumMe dataset.** We used the original title of a video whenever possible. “(no change)” means we used the title as provided in the SumMe dataset. For those videos that we were able to find from the YouTube website, we also include their Youtube unique video identifiers. For the video ID u9E9HpuJQ7U, we changed the title to “Hibachi Cooking” because the original title “Amazing Cooking!”) was not very descriptive of the video content.

Original Title	Substituted Title	Youtube ID (link)
Air Force One	Barack Obama Berlin Die Air Force One schwebt ber TXL ein	uWlftWgKFc8
Base Jumping	Base Jumping, Wingsuit in Norway	8rmIDJEYjoU
Bearpark Climbing	Brenpark Bern Berna, Ursina Bjrk am klettern im Baum	on32amsEoNA
Bike Polo	(no change)	
Bus in Rock Tunnel	Charter bus in rock tunnel	bBcsQcOyuIo
Car over camera	Car runs over camera	
Car railcrossing	Meanwhile At The Railway Crossing	R_0_o1N9hBw
Cockpit Landing	Cockpit landing	
Cooking	Hibachi Cooking	u9E9HpuJQ7U
Eiffel Tower	(no change)	
Excavators river crossing	Crazy Russians crossing river excavators	nGMxQfJe7OA
Fire Domino	Amazing Fire Domino	U1oT_6HzzVI
Jump	Giant Waterslide Jump	w4vW3fsJ8pk
Kids playing in leaves	(no change)	
Notre Dame	(no change)	
Paintball	Paintball fight	
Paluma Jump	Rock Jumping at little Crystal Creek	
Playing ball	Crow and dog playing with pingpongball	ZOYuB7-KHeU
Playing on water slide	(no change)	pxnRYQnz6Ws
Saving Dolphins	Dolphins stranding and incredibly saved	ekmMD8oYU0
Scuba	Scuba diving	
St Maarten Landing	St. Maarten KLM Boeing 747 landing	SCIJ0F62og4
Statue of Liberty	(no change)	
Uncut evening flight	(no change)	Eqs328fkiaQ
Valparaiso Downhill	(no change)	9hvfYvqS-bE

[VT] Changing a vehicle tire				
How to change tires for off road vehicles	How to use a tyre repair kit - Which? guide	Flat tire	KODA Tips How to Repair Your Tyre	When to Replace Your Tires GMC
				
AwmHb44.ouw (5:54)	98MoyGZKHxc (3:07)	J0nA4VgnoCo (9:44)	gzDbas1RIg (4:48)	XzYM3PfTM4w (1:51)
how-to	how-to	vlog	how-to	how-to

[VU] Getting a vehicle unstuck				
The stuck truck of Mark, The rut that filled...	BBC - Train crash	Girl gets van stuck in the back forty	Smart Electric Vehicle Balances on Two Wheels	Electric cars making earth more green
				
HT5vyqe0Xaw (5:22)	sTEELN-vY30 (2:29)	vdmoEJ5YbrQ (5:29)	xwqBXPGE9pQ (3:53)	akI8YFjEmUw (2:13)
egocentric	news	egocentric	interview	news

[GA] Grooming an animal				
Pet Joy Spa Grooming Services — Brentwood, CA	I am a puppy dog groomer	Nail clipper Gloria Pets professional grooming	Dog Grooming in Buenos Aires	How to Clean Your Dog's Ears - Vetoquinol USA
				
i3wAGJaaktw (2:36)	Bhxx-O1Y7Ho (7:30)	0tmA_C6XwfM (2:21)	3eYKfiOEJNs (3:14)	xxdtq8mxegs (2:24)
commercial	vlog	how-to	story	how-to

[MS] Making a sandwich				
Mexican Fried Chicken Sandwich Recipe	Reuben Sandwich with Corned Beef & Sauerkraut	Poor Man's Meals: Spicy Sausage Sandwich	Saigon Sandwich - Vietnamese Sandwiches...	...Joseph Leonard's Fried Chicken Sandwich...
				
WG0MBPpPC6I (6:37)	HI-...g2gn_A (4:03)	Yi4Ij2NM7U4 (6:45)	37rzWOQsNIw (3:11)	LRw_obCPUt0 (4:20)
how-to	how-to	how-to	vlog	story


























[PK] Parkour				
David Belle Fondateur du parkour...	Charlotte Parkour Charlotte Video Project	Singapore Parkour Free Running...	Parkour Camp Leipzig	Jam Parkour Via del Mar
				
cjibtmSLxQ4 (10:47)	b626MiF1ew4 (3:55)	XkqCExn6.Us (3:08)	GsAD1KT1xo8 (2:25)	PJrm840pAUI (4:34)
story	interview	UGC	UGC	UGC

Figure 2. **TVSum50 Dataset (page 1 of 2)**. Videos are grouped into query categories (e.g., [VT] Changing a vehicle tire). For each video, we show its title, thumbnail image, YouTube unique video identifier, duration, and genre. Click thumbnail images to watch videos.

[PR] Parade				
Chinese New Year Parade NYC Chinatown	Alumnae Parade & Laurel Chain Ceremony	LA Kings Stanley Cup South Bay Parade	Chinatown Parade	Southend-on-Sea and Kortrijk Bike Friendly...
				
91IHQYk1IQM (1:50)	RBCABdtQmI (6:04)	z_6gVvQb2d0 (4:36)	fWutDQy1nnY (9:45)	4wU_LUjG5Ic (2:47)
UGC	UGC	egocentric	vlog	UGC

[FM] Flash mob gathering				
Flash mob protest for Syria Sydney, Australia	ICC World Twenty Bangladesh Flash Mob...	Bluffton Teachers Flash Mob Dance...	ICC World Twenty Bangladesh, Flash Mob...	CASACL - Flashmob in Copenhagen underground...
				
VuWGsYPqAX8 (3:36)	JKpqYvAdIsw (2:32)	xmEERLqJ2kU (7:26)	byxOvuiIJV0 (2:34)	_xMr-HKMfVA ()
UGC	UGC	egocentric	UGC	UGC

[BK] Beekeeping				
Beekeeping 101	A Year of Beekeeping	Paper Wasp Removal From the Ground Up	Killer Bees Kill 1000-lb Hog in Bisbee AZ	Apis Mellifera in a Vertical Log Hive
				
WxtbjNsCQ8A (4:25)	uGu_10sucQo (2:47)	EE-bNr36nyA (1:38)	Se3oxnaPsz0 (2:18)	oDXZc0tZe04 (6:20)
interview	UGC	how-to	news	UGC

[BT] Attempting a bike trick				
...How to Whip a Motocross Bike	Smage Bros. Motorcycle Stunt Show	How to stop your Fixie	How to lock your bike. The RIGHT way!	Pure Fix TV: How to Wheelie
				
qqR6AEXwxoQ (4:29)	EYqVtI9YWJA (3:18)	eQu1rNs0an0 (2:44)	JgHubY5Vw3Y (2:23)	iVt07TCKFM0 (1:44)
interview	interview	how-to	how-to	how-to






[DS] Dog show				
German Shepherd Dog Show (KCI)...	The Dog Show HD	Oliver's Show - Dog's tale	...Obie the obese dog works toward weight loss	Will A Cat Eat Dog Food?
				
E11zDS9XGzg (8:30)	NyBmCxDoHJU (3:09)	kLxoNp-UchI (2:10)	jcoYJXDG9sw (3:19)	-esJrBWj2d8 (3:50)
vlog	UGC	UGC	news story	egocentric

Figure 3. TVSum50 Dataset (page 2 of 2). Videos are grouped into query categories (e.g., [PR] Parade). For each video, we show its title, thumbnail image, YouTube unique video identifier, duration, and genre. Click thumbnail images to watch videos.

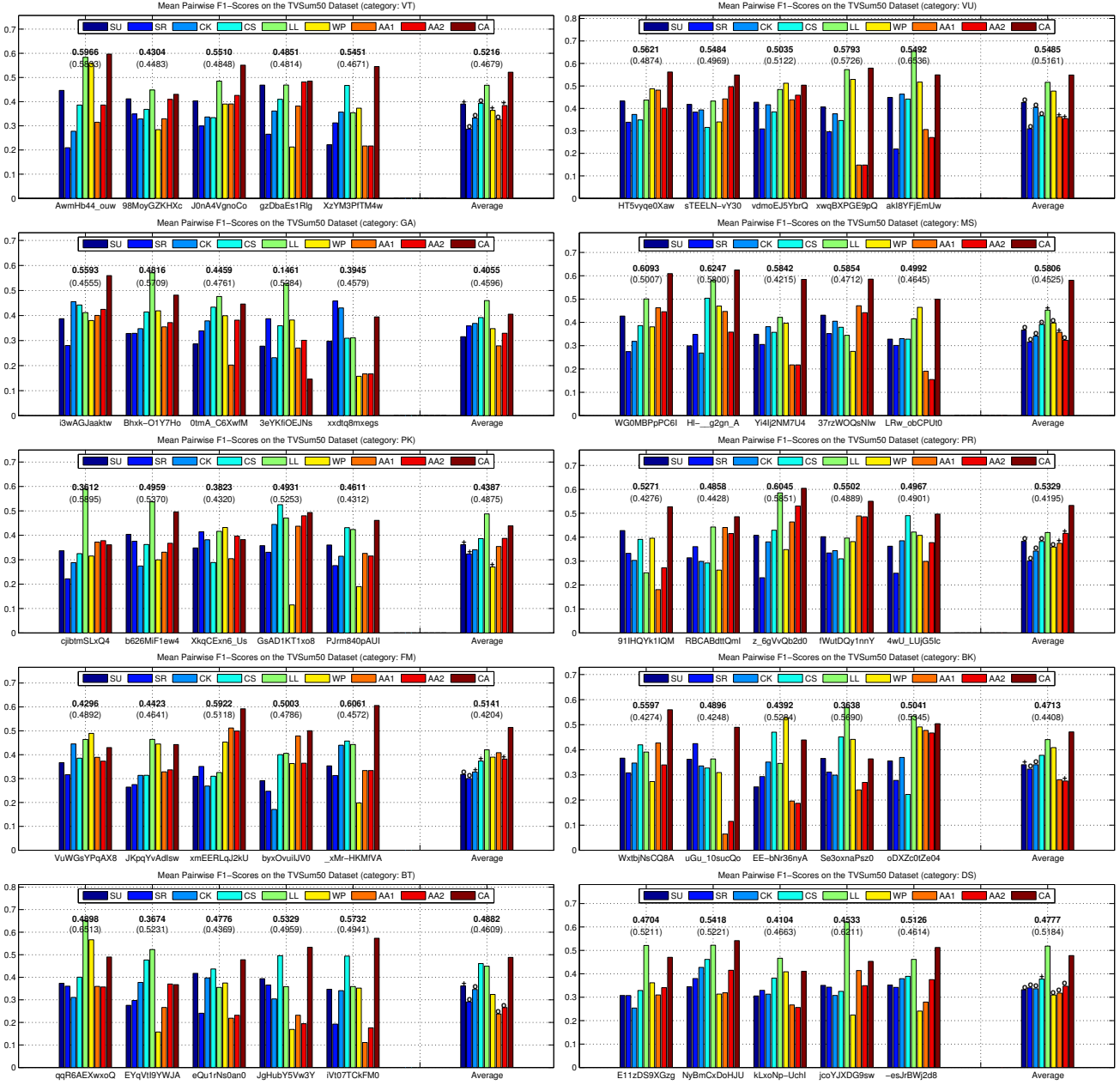


Figure 4. **Experimental results on each video category on the TVSum50 dataset.** Models compared include uniform sampling (SU), random sampling (SR), k -means clustering (CK), spectral clustering (CS), LiveLight by Zhao and Xing (LL), web image prior by Khosla *et al.* (WP), archetypal analysis with video frames (AA₁), archetypal analysis with a combination of video frames and images (AA₂), and our co-archetypal analysis (CA). Bold-faced numbers represent our results; numbers in parentheses represent the highest score from any baseline. The last column shows average scores, where we also show the statistical significance of the pairwise difference between our approach (CA) and the baselines (the 'o' mark indicates $p < 0.01$; the '+' mark indicates $p < 0.05$). Except for a few videos, our approach outperforms all other baselines.

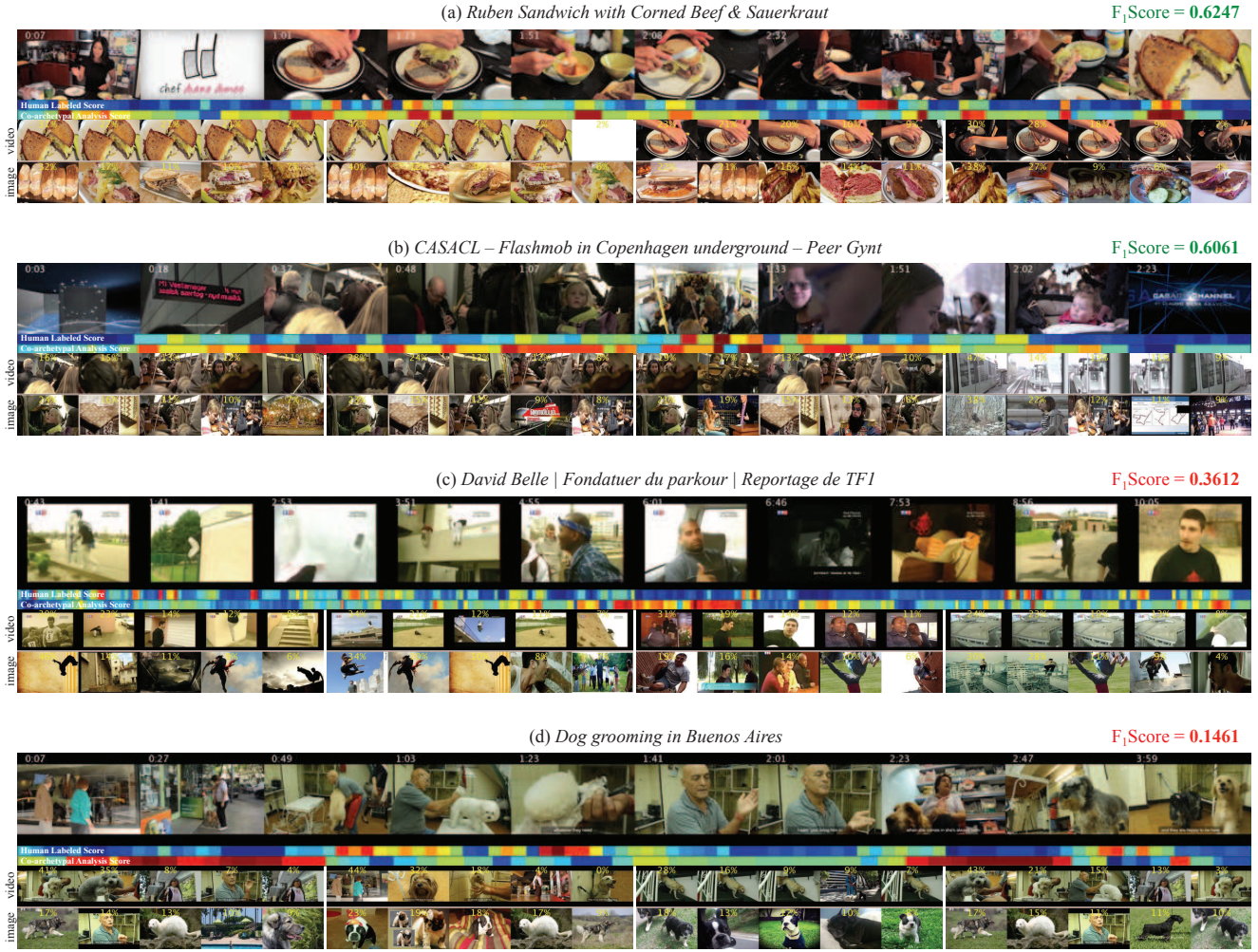


Figure 5. **Detailed results of co-archetypal analysis on the TVSum50 dataset.** Shown here are four example results (a-d), each with a video title, an F_1 score of our approach, video frames, two heatmap representations of normalized shot importance scores (one from human labels, another from our approach), and four learned co-archetypes represented in terms of video frames and images; stacked in that order. (a-b) Two example results with highest F_1 scores; (c-d) two example results with lowest F_1 scores. The results with high F_1 scores tend to show co-archetypes that match more closely with video segments scored higher by humans (e.g., (a) sandwich, (b) underground) compared to the results with low F_1 scores.