

Automatic Construction of an Action Video Shot Database using Web Videos



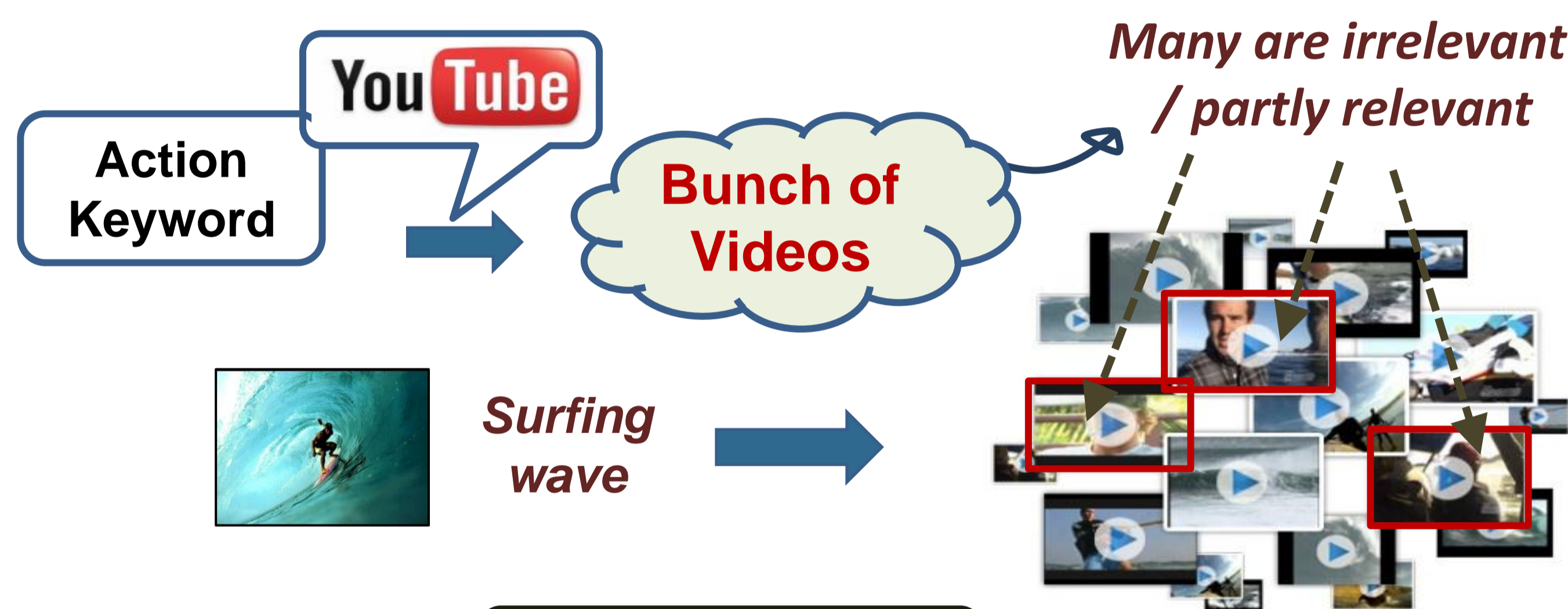
The University of Electro-Communications, Tokyo, Japan

DO HANG NGA and KEIJI YANAI

Summary

- A novel method to collect “Action” video shots from the Web
 - Fully-automatic and unsupervised
 - Only providing “Action keywords” such as “paint+picture” at first
 - Tag-based video selection and ST-feature-based shot selection
- Large-scale experiments for 100 kinds of human actions
 - 36.6% prec@100 for 100 actions, 79.8% for the top 10 actions

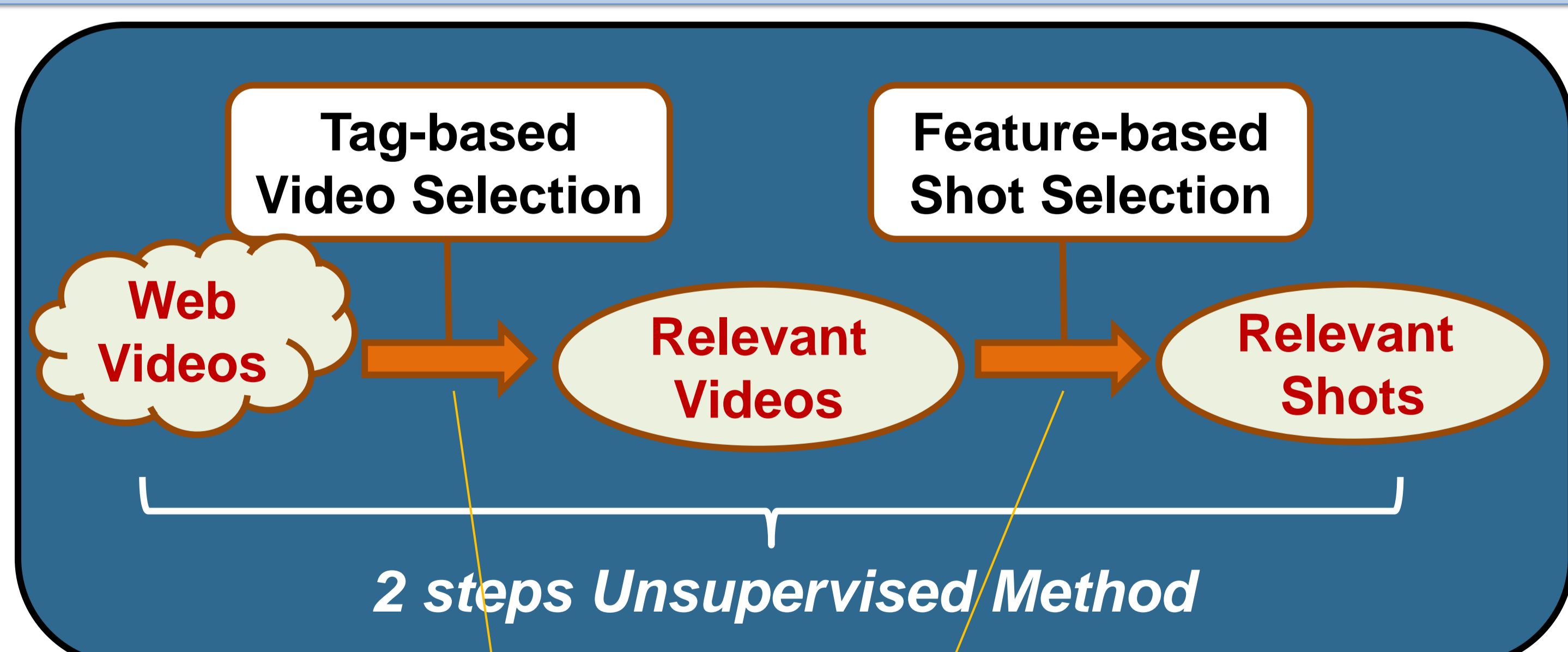
Introduction & Objective



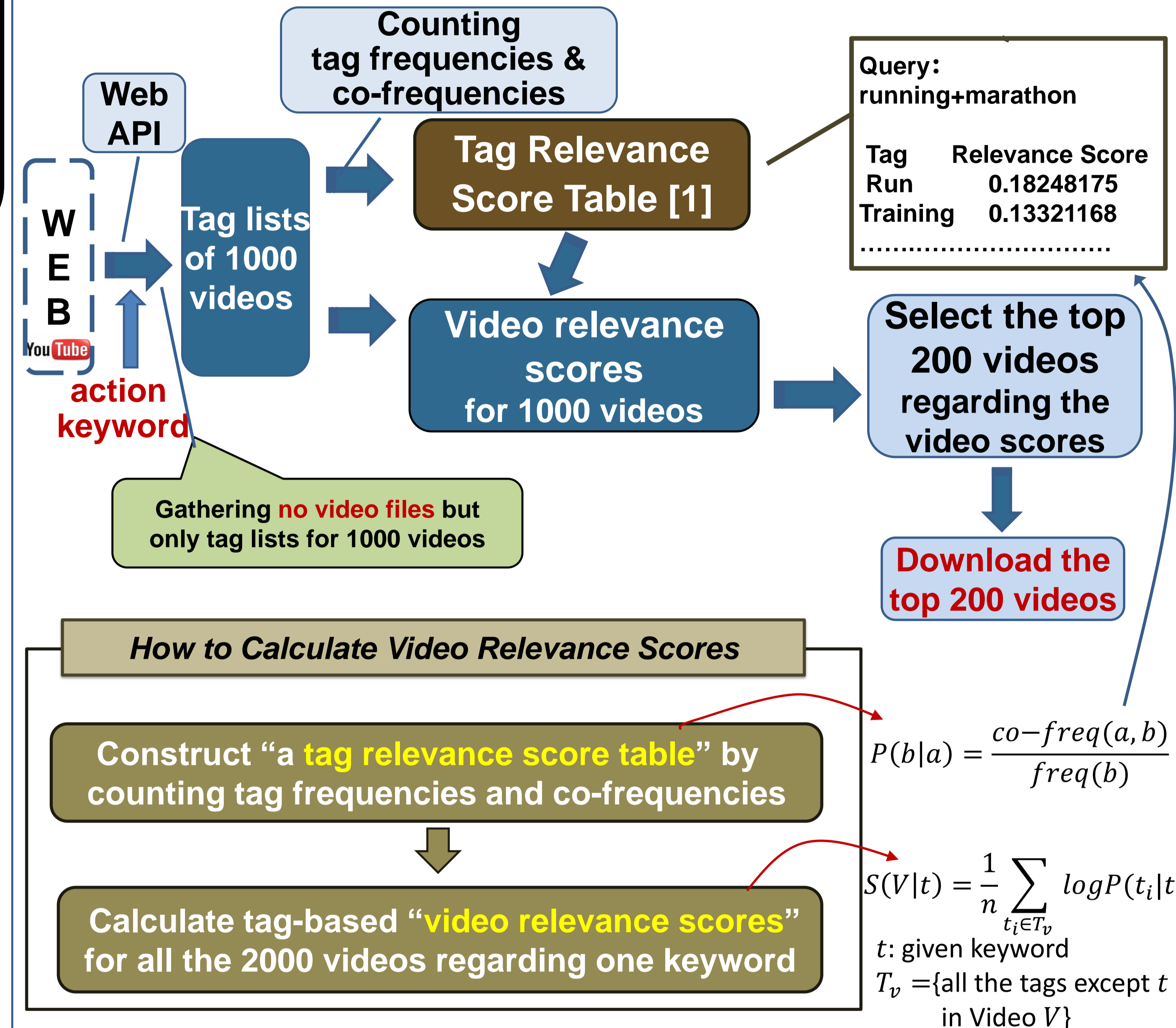
Objective



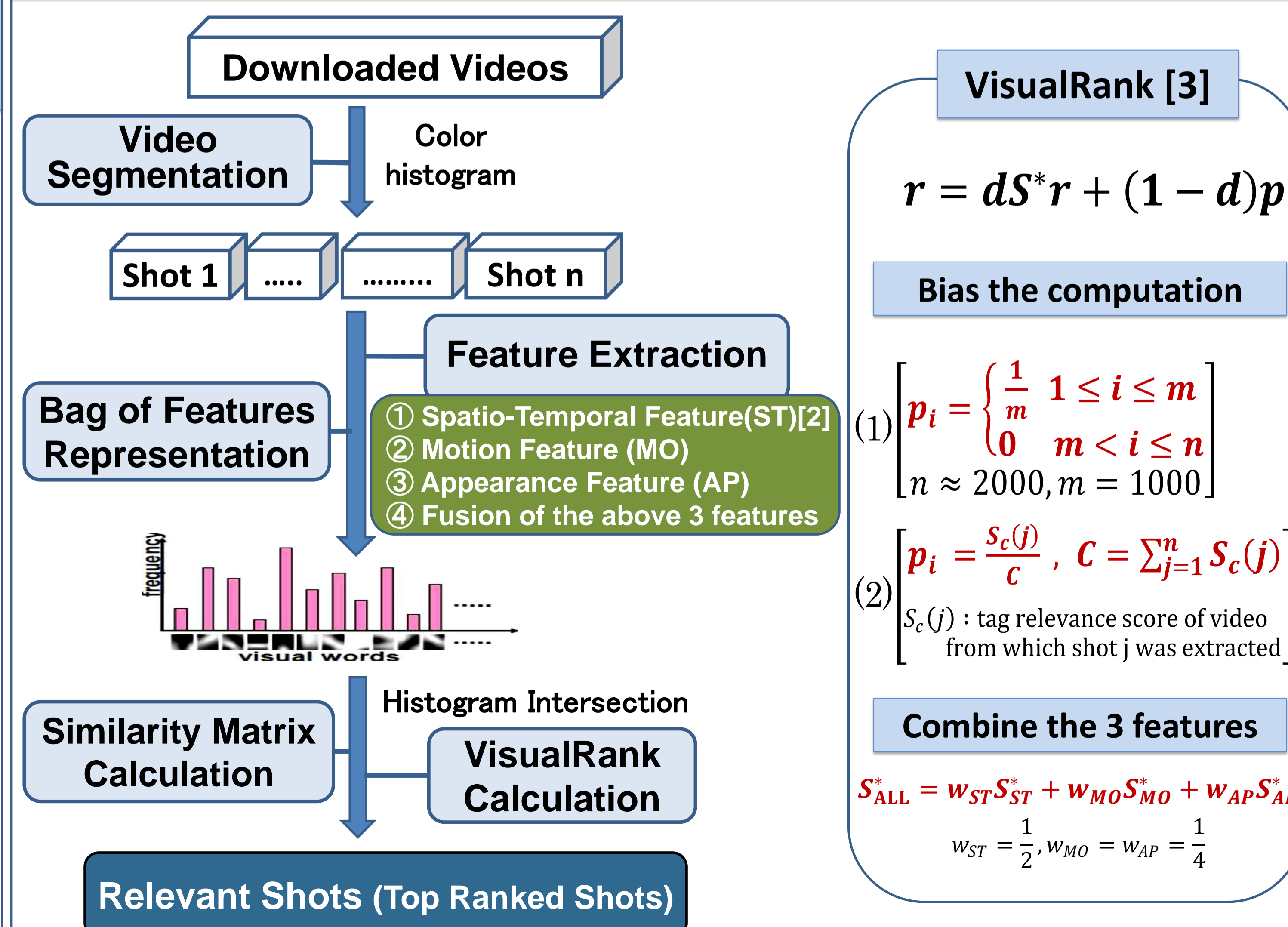
Overview of Proposed Method



Tag-based Video Selection



Feature-based Shot Selection



Experiments & Results

(1) Six-action experiments under various conditions



Table 1: The conditions of the experiments and the results (Mean Precision@100 of 6 actions)

Exp No.	Tag-based Ranking	Biased damp. vec.	Visual Feature	Mean prec@100
RND	Randomly-selected 100 shots			14.2%
TAG	✓	—	—	23.5%
1	—	—	ST	33.7%
2	✓	—	ST	41.0%
3(1)	✓	✓(1)	ST	47.3%
3(2)	✓	✓(2)	ST	44.8%
5	✓	✓(1)	Motion	31.8%
6	✓	✓(1)	Appear.	39.7%
7	✓	✓(1)	Fusion	49.5%

(2) Large-scale 100-action experiments

Table 2: Prec@100 of Top 60 in 100 human actions Results (%)

Action	Prec	Action	Prec	Action	Prec
soccer+dribble	100	shoot+football	58	shoot+football	33
fold+origami	96	tie+shoelace	57	draw+eyebrows	32
crochet+hat	95	laugh	50	fieldhockey+dribble	32
arrange+flower	94	dive+sea	49	hit+golfball	32
paint+picture	88	harvest+rice	49	lunge	32
boxing	86	ski	49	play+piano	32
comb+hair	83	iron+clothes	47	row+boat	32
parachute+jump	82	twist+crunch	47	sing	32
do+exercise	79	dance+flamenco	45	chat+friend	31
do+aerobics	78	dance+hiphop	43	clean+floor	31
do+yoga	77	eat+ramen	42	cut+onion	31
serve+volleyball	75	dance+tango	41	shave+mustache	31
surf+wave	75	play+trumpet	41	pick+lock	30
shoot+arrow	73	play+drum	40	plaster+wall	30
fix+tire	67	skate	37	blow+candle	29
basketball+dribble	64	swim+crawl	36	wash+face	29
blow-dry+hair	64	cut+hair	35	walking+street	29
ride+bicycle	62	run+marathon	35	brush+teeth	28
bowl+ball	58	count+money	33	catch+fish	28
curl+bicep	58	paint+wall	33	drive+car	28

See 100 action shot results at <http://mm.cs.uec.ac.jp/webvideo/>

- [1] Q.Yang, X.Chen and G.Wang: “We2.0 Dictionary”, Proc. of ACM International Conference on Image and Video Retrieval (CIVR), p.591–600, 2008.
- [2] A.Noguchi and K.Yanai: “A SURF-based Spatio-Temporal Feature for feature-fusion-based action recognition”, Proc. Of ECCV WS on Human Motion: Understanding, Modeling, Capture and Animation, 2010.
- [3] Y. Jing and S. Baluja: “Visualrank: Applying pagerank to large-scale image search”, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 30, No. 11, pp. 1870-1890, 2008.