ACM MM'13

Large-scale Web Video Shot Ranking Based on Visual Features and Tag Co-occurrence

The University of Electro-Communications **Do Hang Nga** **Keiji Yanai**

Objective

Extracting automatically relevant video shots of specific actions from Web videos

Overview of Proposed system

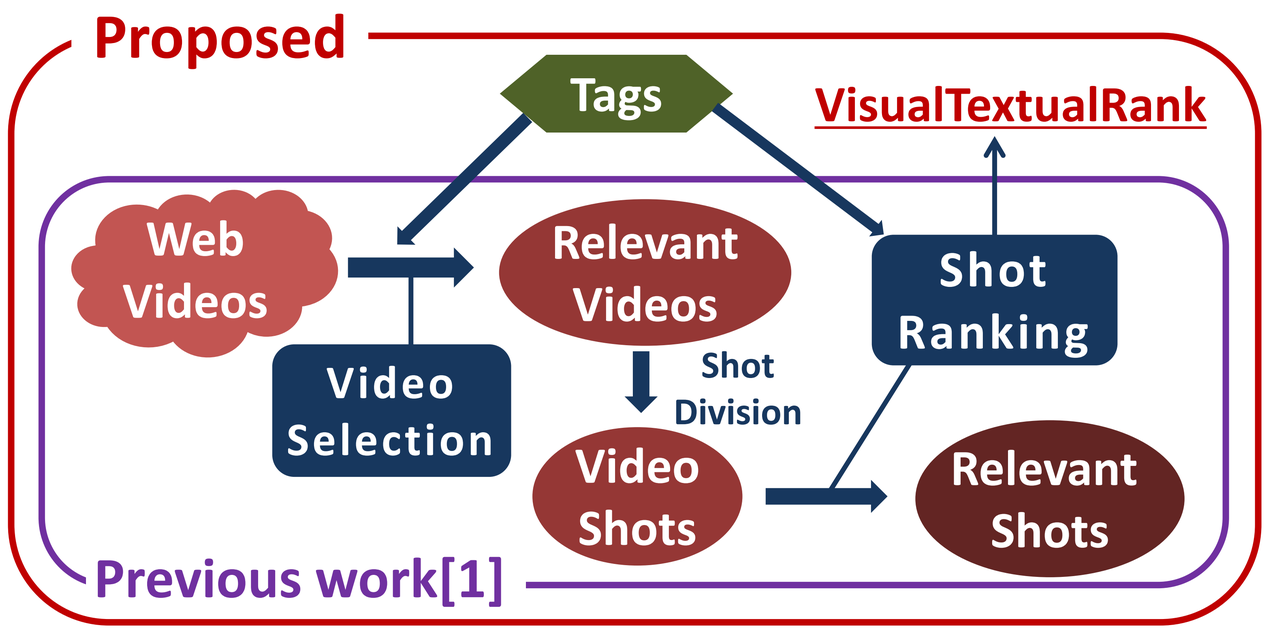
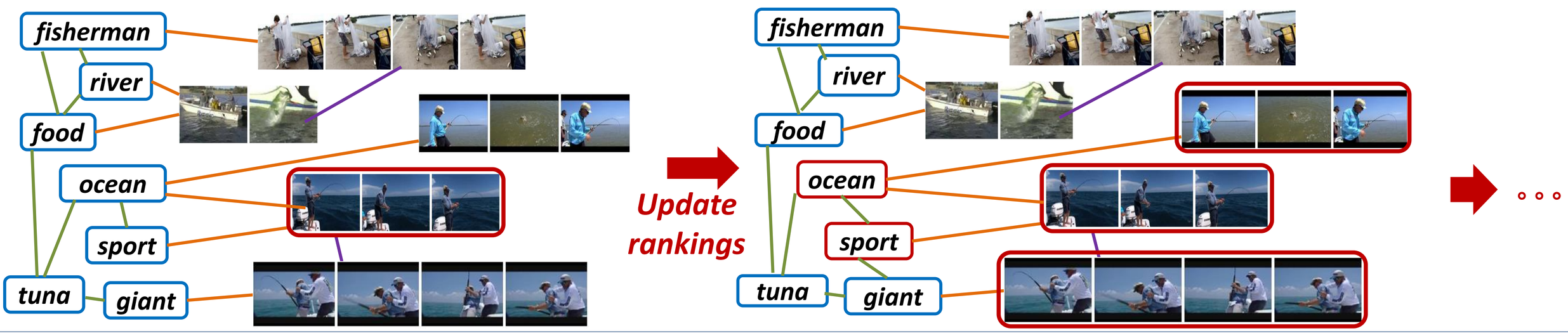


Illustration of VisualTextualRank

Iterative shot-tag co-ranking based on shot-tag linkage and visual features of shots



Previous work

- **Tag co-occurence based Video Selection:**

$$Sc(V/t) = \frac{1}{|\mathcal{T}_V|} \sum_{t_i \epsilon \mathcal{T}_V} log_2 F(t, t_i)$$

$F(t, t_i)$: *number of videos tagged with both t and* t_i
$\mathcal{T}_V$: *tag set of video V*

- **VisualRank[3] based Shot Ranking:**

$$r = dS^* r + (1-d)p \qquad p_i = \begin{cases} \frac{1}{k} & , i \le k \\ 0 & , i > k \end{cases}$$

Rank Vector · Similarity Matrix · Damping Vector

S_{ij} = *histogram intersection between spatio − temporal features of shot i and shot j*

Proposed Shot Ranking Method

$$RS_k = \alpha \times SM^* \times SC^* \times RT_k + (1-\alpha)p$$
$$RT_{k+1} = (SC')^* \times RS_k$$

RS: *Rank vector of shots* RT: *Rank vector of tags*
SM: *shot- shot similarity matrix* SC: *shot- tag similarity matrix*

✓ **Explore the mutual reinforcement between shots and textual information of their videos to improve shot ranking**

Experiments and Results

- **Data: 100 actions (Previous work[1])**
- **Comparison between VR[1] and VTR(proposed)**
- **Evaluation Method: Prec@100**

Action	VR	VTR	Action	VR	VTR	Action	VR	VTR	Action	VR	VTR	Action	VR	VTR	Action	VR	VTR	Action	VR	VTR
Soccer+dribble	100	100	Batting	66	61	Dance+hiphop	43	68	Lunge	32	27	Drive+car	28	40	Row+dumbell	16	30	Swim+butterfly	7	9
Fold+origami	96	99	Basketball+dribble	64	87	Eat+ramen	42	47	Play+piano	32	34	Plant+flower	28	24	Wash+clothes	15	29	Bake+bread	6	8
Crochet+hat	95	97	Blow-dry+hair	64	59	Dance+tango	41	41	Row+boat	32	23	Play+guitar	28	41	Wash+dishes	15	39	Cook+rice	6	11
Arrange+flower	94	96	Knit+sweater	64	68	Play+trumpet	41	59	Sing	32	65	Lift+weight	27	51	Comb+hair	14	26	Grill+fish	5	13
Paint+picture	88	87	Ride+bicycle	62	70	Play+drum	40	45	Chat+friend	32	52	Raise+leg	27	40	Drink+coffee	14	16	Jog	5	6
Boxing	86	84	Curl+bicep	58	59	Skate	37	42	Clean+floor	31	38	Hang+wallpaper	26	46	Swim+breaststroke	13	18	Slice+apple	5	16
Jump+parachute	82	83	Shoot+ball	58	58	Swim+crawl	36	49	Cut+onion	31	24	Jump+rope	26	49	Cry	12	12	Peel+apple	5	14
Jump+trampoline	82	92	Tie+shoelace	57	73	Cut+hair	35	42	Shave+mustache	31	30	Climb+tree	24	24	Eat+sushi	12	23	Bowl+ball	4	4
Do+exercise	79	61	Laugh	50	54	Run+marathon	35	43	Pick+lock	30	28	Ride+horse	24	15	Serve+tennis	11	27	Smile	4	6
Do+aerobics	78	79	Dive+sea	49	41	count+money	33	58	Plaster+wall	30	55	Roll+makizushi	24	36	Tie+necktie	11	28	Kiss	2	3
Do+yoga	77	70	Harvest+rice	49	46	Paint+wall	33	32	Blow+candle	29	44	Sew+button	24	46	Boil+egg	9	11	AVG.	36.6	43.5
Surf+wave	75	73	Ski	49	60	Shoot+football	33	29	Wash+face	29	24	Fry+tempura	23	12	Head+ball	9	16			
Shoot+arrow	73	81	Iron+clothes	47	48	Draw+eyebrows	32	32	Walking+street	29	46	Slap+face	20	45	Swim+backstroke	9	9			
Massage+leg	72	78	Twist+crunch	47	32	Fieldhockey+dirbble	32	68	Brush+teeth	28	27	Read+book	19	21	Take+medicine	8	7			
Fix+tire	67	77	Dance+flamenco	45	53	Hit+golfball	32	70	Catch+fish	28	59	Squat	19	34	Serve+volleyball	7	40			

- **A novel shot ranking method using visual features and tag co-occurence**
- **Obtained better results than baseline which uses only visual features**

[1] Do Hang Nga and Keiji Yanai: Automatic Construction of an Action Video Shot Database using Web Videos. ICCV2011.
[2] Do Hang Nga and Keiji Yanai: Automatic Collection of Web Video Shots Corresponding to Specific Actions using Web Images. LSVSM'12
[3] Y. Jing and S. Baluja. Visualrank: Applying pagerank to large-scale image search. PAMI, Vol. 30, No. 11, pp. 1870-1890, 2008.