

#### Binary Coding for Partial Action Analysis with Limited Observation Ratios Jie Qin<sup>1,2</sup>, Li Liu<sup>3,4</sup>, Ling Shao<sup>4</sup>, Bingbing Ni<sup>5</sup>, Chen Chen<sup>6</sup>, Fumin Shen<sup>7</sup> and Yunhong Wang<sup>1</sup> <sup>2</sup>ETH Zurich <sup>3</sup>Malong Technologies Co., Ltd. <sup>1</sup>Beihang University <sup>4</sup>University of East Anglia <sup>6</sup>University of Central Florida <sup>7</sup>University of Electronic Science and Technology of China <sup>5</sup>Shanghai Jiao Tong University

# **Motivation:**

- Most action recognition approaches analyze after-the-fact actions. However, capturing complete actions is often difficult due to occlusions, interruptions, etc.
- Partial action recognition (PAR) has a wide range of applications in *intelligent surveillance, smart homes, retrieval systems, etc.*
- $\succ$  Existing PAR methods are:
- Lack of generality, e.g., requiring sufficient observations, partial observations from the beginning, known observation ratios (ORs);
- Lack of scalability, i.e., developed based on high-dimensional video data, involving unacceptable memory usage and computational costs for real-time applications.

**Traditional Action Recognition** 

**Action Prediction** 

Partial Action Recognition (Ours)

# **Contributions:**

- (1) Perform partial action analysis in a more general and practical scenario, where a *short* temporal video segment of *unknown* ORs, observed during *any* period of the complete execution is utilized for action analysis.
- (2) Propose a joint learning framework, which collaboratively addresses feature reconstruction and binary coding, based on discrete alternating optimization.
- (3) Present our approach in both supervised and unsupervised fashions and systematically evaluate it on four action benchmarks in terms of three tasks, i.e., *partial action retrieval*, recognition and prediction.







Modelling the fitting error

**Discrete constraint** 

$$\mathbf{x}_i^M, \dots, \mathbf{x}_N^1, \dots, \mathbf{x}_N^M]$$

$$[\underbrace{\mathbf{y}_i}_{i}, \ldots, \underbrace{\mathbf{y}_N, \ldots, \mathbf{y}_N}_{M \text{ times}}]$$

$$||\mathbf{y}_i - \mathbf{W}^{\mathrm{T}} \mathbf{x}_i^m||_2^2$$

$$= \operatorname{sign}(\mathbf{P}^{\mathrm{T}}\mathrm{g}(\mathbf{x}))$$

$$bel_j^n,$$
  
 $bel_j^n,$ 

Discrete alternating optimization without any relaxation

# **Experiments:**

### > Partial action retrieval:

Method			MAP (%)	Precision @radius2 (%)	Precision @rank50(%)	Training time (s)	Test coding time (s)		0.5
Single-Modality Binary Coding Methods		SDH [46]	29.80	0.11	37.92	477.17	$3.8 \times 10^{-6}$	0.4	
	Supervised	FastHash [21]	47.46	0.001	54.05	$1.56 \times 10^{3}$	$9.3 \times 10^{-4}$		
		KSH [34]	5.10	0.095	8.77	$1.17 \times 10^{3}$	$3.4 \times 10^{-6}$		
		CCA-ITQ [8]	34.71	2.58	42.61	8.90	$2.9 \times 10^{-6}$		₽ 0.2
	Unsupervised	AGH [33]	3.08	1.27	2.10	35.58	$3.8 \times 10^{-6}$	0.1	0.1
		PCA-ITQ [8]	2.94	< 0.001	2.44	7.62	$4.4 \times 10^{-6}$		
Cross-Modality Binary Coding Methods	Supervised	SePH [22]	50.20	23.42	54.18	$2.14 \times 10^{3}$	$3.7 \times 10^{-6}$	32 48 64 80 96 128 Code length	
		SCM [58]	37.14	3.11	43.62	204.52	$8.5 \times 10^{-6}$	0.9	0.7
		CVH [17]	14.41	1.54	24.98	25.21	$2.2 \times 10^{-6}$		0.6
		CMSSH [2]	35.87	1.85	37.85	895.75	$6.4 \times 10^{-6}$	0.7	CSD 0.5
	Unsupervised	CMFH [6]	5.02	2.42	3.93	411.37	$7.3 \times 10^{-6}$		
Proposed	Supervised	PRBC-Sup	<b>59.71</b> ±0.754	<b>32.31</b> ±0.521	<b>63.24</b> ±0.630	129.01	$3.4 \times 10^{-6}$		0.3
	Unsupervised	PRBC-Unsup	32.27±0.717	$16.94 \pm 0.448$	39.80±0.692	144.34	$3.2 \times 10^{-6}$	0.3	
4096-d C3D Feature (CF)			2.91	-	2.01	-	-	0.2	
4096-d C3D Feature+Reconstruction (CF+R)			12.4	-	10.76	129.01	-		
								32 48 64 80 96 128	32 48 64

### Partial action recognition:

	<b>16-frame</b> partial actions for testing						<b>32-frame</b> partial actions for testing						
Method	HMDB51			UCF101			HMDB51			UCF101			
		32 bits	64 bits	128 bits	32 bits	64 bits	128 bits	32 bits	64 bits	128 bits	32 bits	64 bits	128 bits
	SDH [46]	13.91	16.47	19.36	27.63	38.05	44.78	12.31	15.15	19.35	33.06	42.78	50.33
Single-Modality Binary Coding Methods	FastHash [21]	16.70	21.08	23.09	37.17	48.57	55.98	15.15	18.14	20.11	39.51	49.31	56.29
	CCA-ITQ [8]	17.45	19.03	21.23	49.23	52.59	54.90	19.42	20.80	22.63	52.53	56.77	59.57
	KSH [34]	2.23	2.85	2.62	2.87	2.28	2.47	2.58	2.73	2.31	7.98	8.02	8.85
	AGH [33]	5.36	6.20	5.5	1.97	1.94	1.47	3.33	3.33	4.62	3.74	4.40	3.82
	PCA-ITQ <sup>8</sup>	2.30	3.08	3.22	4.74	4.94	4.74	4.02	3.63	3.87	6.32	7.39	7.11
	SePH [22]	32.89	33.97	37.15	57.61	63.61	67.84	37.07	39.58	41.24	59.21	65.06	69.11
Cross-Modality	SCM [58]	31.78	36.48	38.67	40.94	62.06	68.57	31.57	35.75	37.95	41.03	62.29	68.97
Binary Coding Methods	CVH [17]	25.52	31.13	34.93	45.07	56.78	64.70	26.32	31.55	36.04	45.90	57.92	66.17
	CMFH 6	2.65	2.60	3.15	7.04	7.32	7.95	3.94	3.85	4.91	8.84	8.42	9.53
Dropogod	PRBC-Sup	42.78	45.80	48.60	70.27	75.11	78.46	46.52	49.32	50.79	71.79	77.47	80.80
r roposed	PRBC-Unsup	29.64	32.76	34.25	58.06	62.94	67.15	31.64	33.90	34.84	56.15	60.49	64.19
	CCA* 9	39.51 (2048-d)			70.61 (4096-d)			4]	.87 (2048	-d)	72.26 (4096-d)		
Cross-View	PLSR* [54]	37.71 (4096-d)			66.83 (4096-d)			40.02 (4096-d)			68.05 (4096-d)		
Feature Learning Methods	XQDA* [20]	11.53 (512-d)			40.11 (512-d)			14.32 (512-d)			44.14 (512-d)		
	CVFL [55]	40.32 (4096-d)			70.97 (4096-d)			44.12 (4096-d)			73.13 (4096-d)		
(a) Full Actions (4096-d C3D feature) (b) Partial Actions (4096-d C3D feature)													

### Action prediction:

Mathad	UT-Inte	eraction da	UT-Interaction d		
Method	OR=0.1	OR=0.2	OR=0.3	OR=0.1	OR=0.2
Bayesian [41]	16.7	16.7	16.7	16.7	16.7
BP-SVM [41]	16.8	21.7	27.8	16.7	24.0
IBoW [41]	14.5	17.9	30.8	16.8	29.9
DBoW [41]	15.2	20.2	30.7	16.7	28.9
SC [3]	18.3	33.3	56.7	21.7	43.3
MSSC [3]	18.3	40.0	60.0	21.7	40.0
MTSSVM [15]	36.7	46.7	66.7	33.3	50.0
RPT [57]	13.3	26.7	56.7	15.0	33.3
AAC [56]	45.0	46.7	60.0	51.3	53.3
MOVEMES [18]	38.3	54.5	68.3	31.3	41.3
MMAPM [14]	46.7	51.7	70.0	36.7	55.0
PRBC-Sup@64bits	55.0	58.3	63.3	60.0	65.0
PRBC-Sup@128bits	56.7	58.3	65.0	60.0	63.3



## Dataset&Feature: HMDB51/UCF101&C3D; UT-Interaction&Cuboids+BoW Partial actions: 16/32-frame segments from full action videos (OR < 30%)





