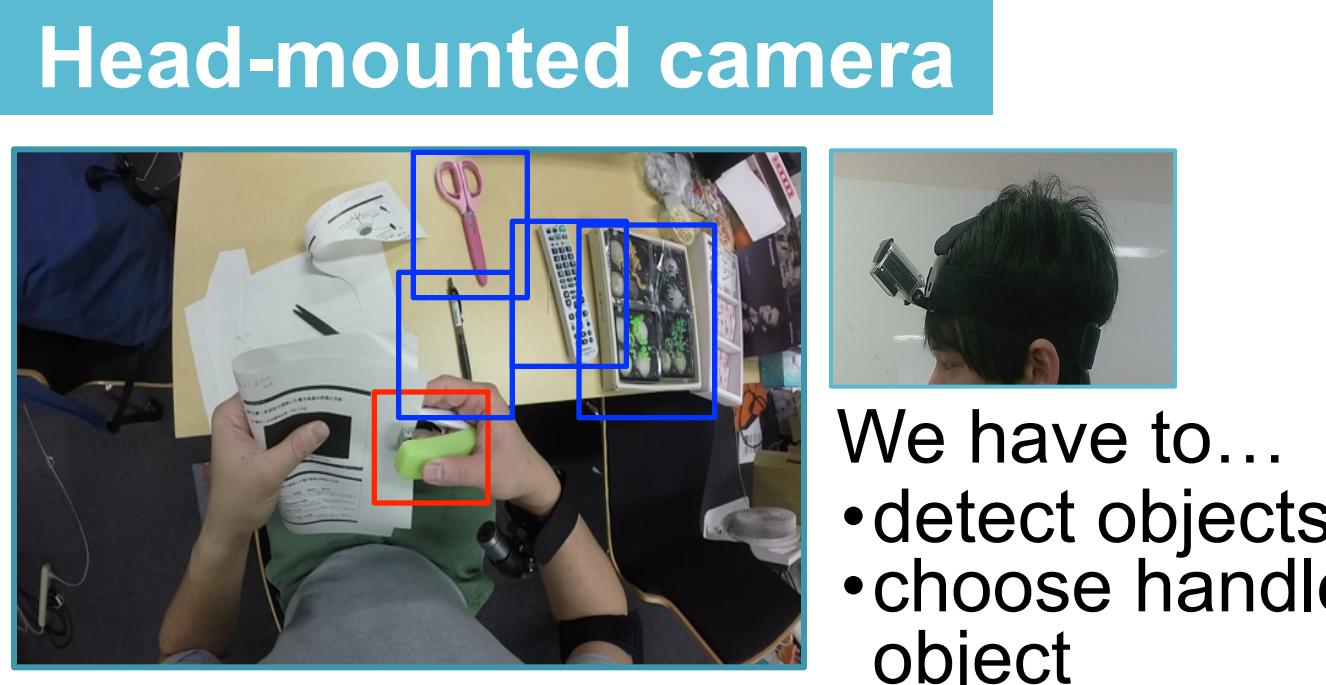


Recognizing Activities of Daily Living with a Wrist-mounted Camera

Katsunori Ohnishi, Atsushi Kanehira, Asako Kaneko, Tatsuya Harada
 The University of Tokyo



Overview



Goal:

Existing works:

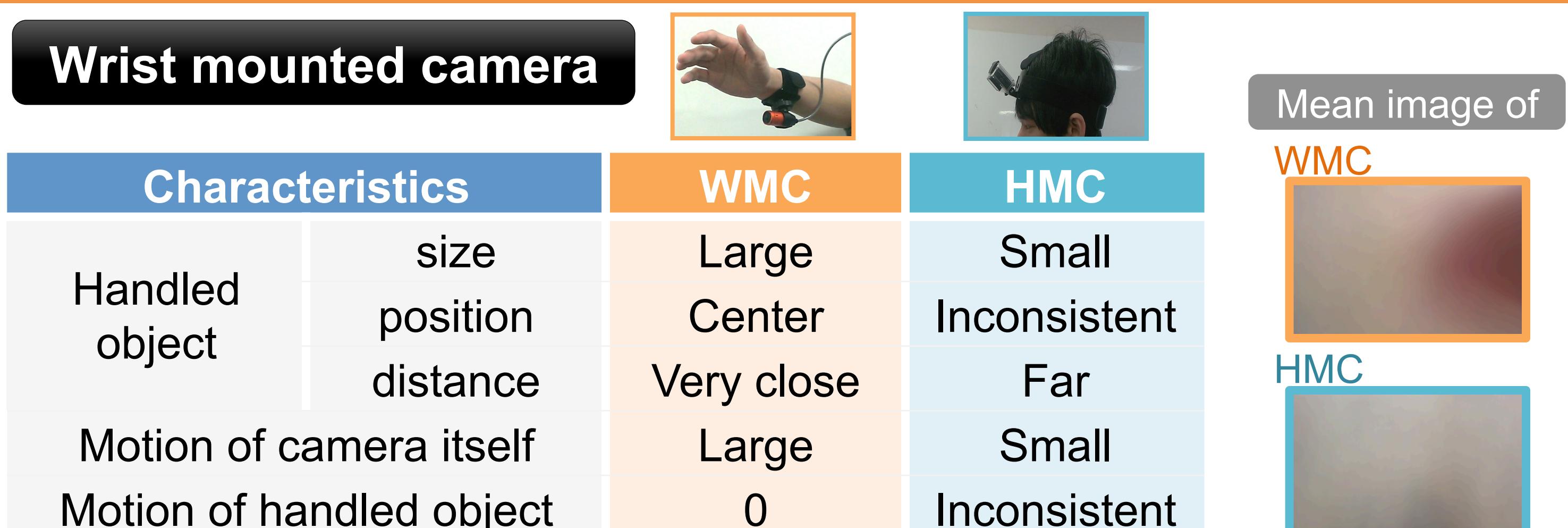
Idea:

Contributions:

Wrist mounted camera



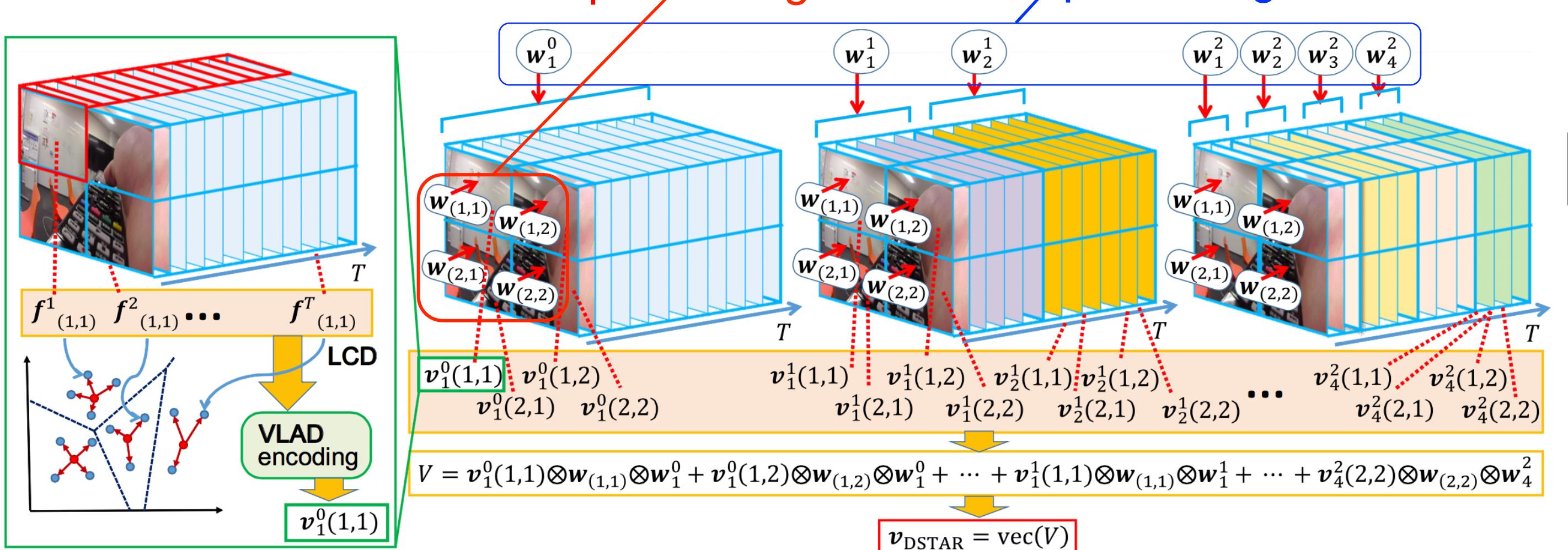
Characteristics	WMC	HMC
Handled object	size position distance	Large Center Very close
Motion of camera itself	Large	Small
Motion of handled object	0	Inconsistent



Algorithm

DSTAR: Discriminative Spatio-Temporal Aggregated latent concept descriptorRs

Illustration of DSTAR



DSAR

DSAR contains spatial information even after VLAD coding

$$V_{\text{DSAR}} = \sum_{i=1}^a \sum_{j=1}^a v(i,j) w_{(i,j)(i,j)}^T$$

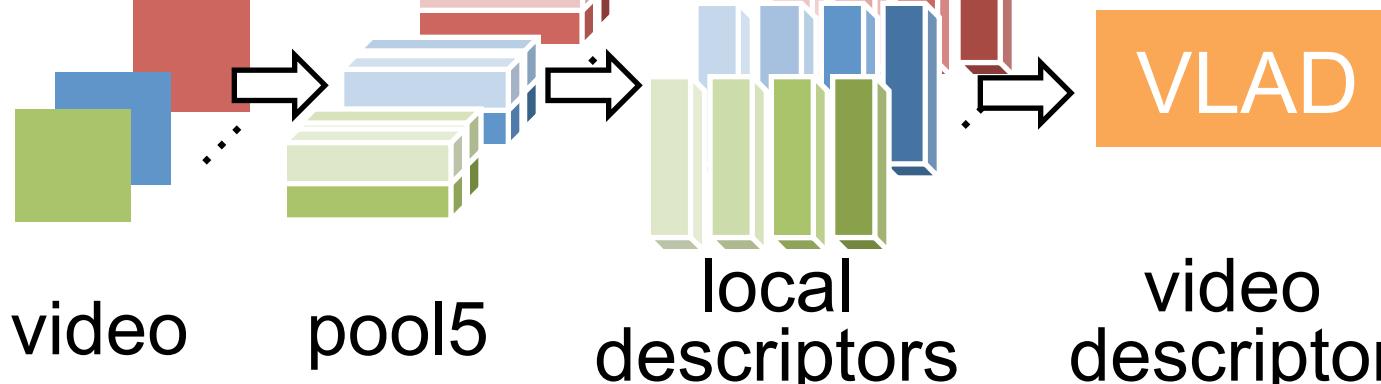
DSTAR

DSTAR contains spatial and temporal information even after VLAD coding

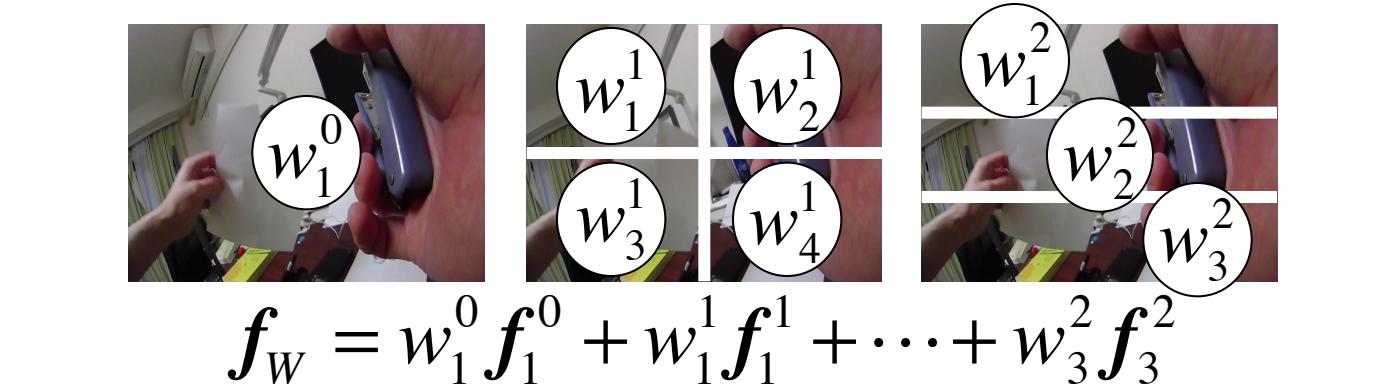
$$V_{\text{DSTAR}} = \sum_{i=1}^a \sum_{j=1}^a \sum_{l=0}^L \sum_{s=1}^{2^l} v_s^l(i,j) \otimes w_{(i,j)} \otimes w_s^l$$

Optimize **spatial weights** and **temporal weights** iteratively and alternately.

Latent Concept Descriptors [Z. Xu, et al., CVPR 2015]



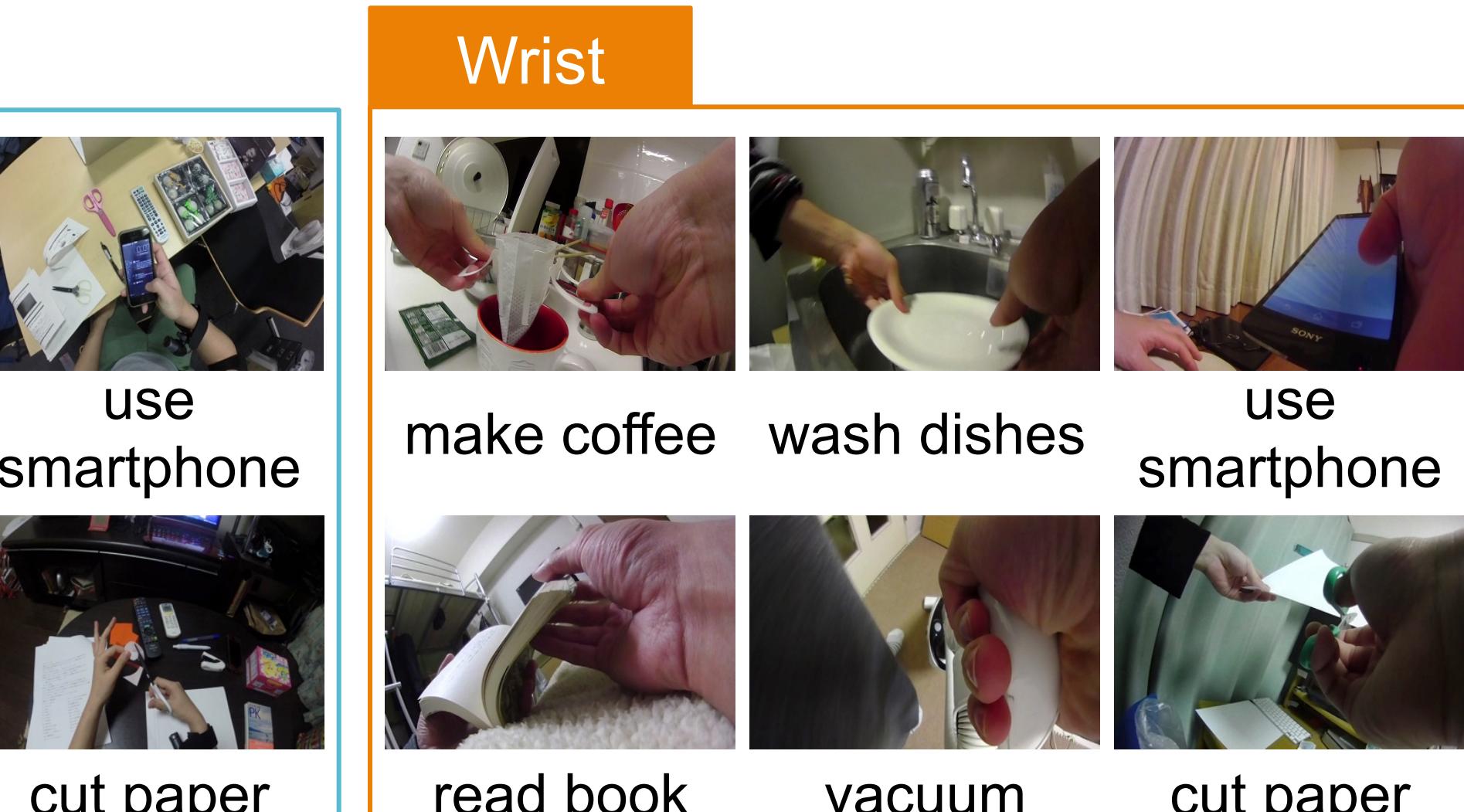
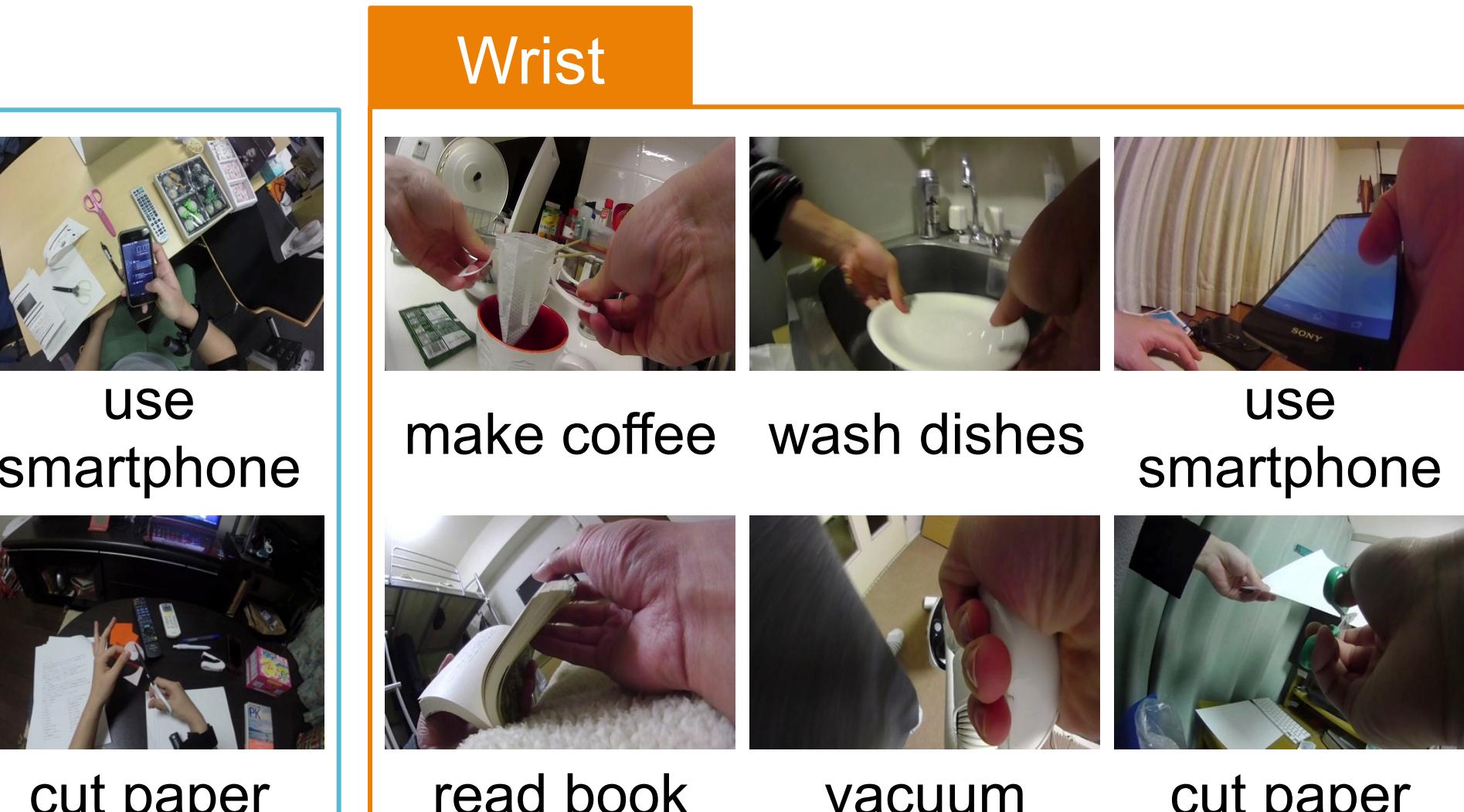
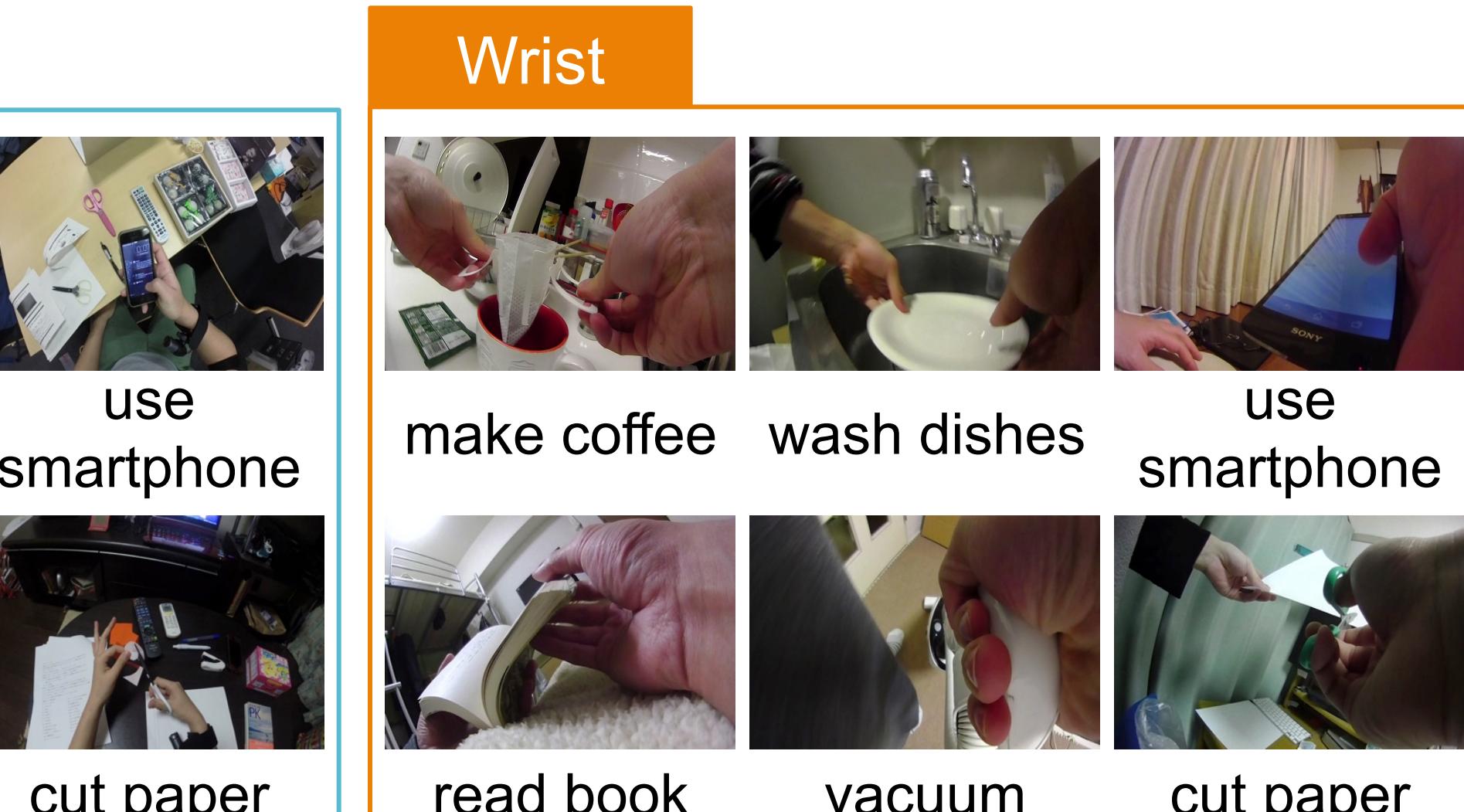
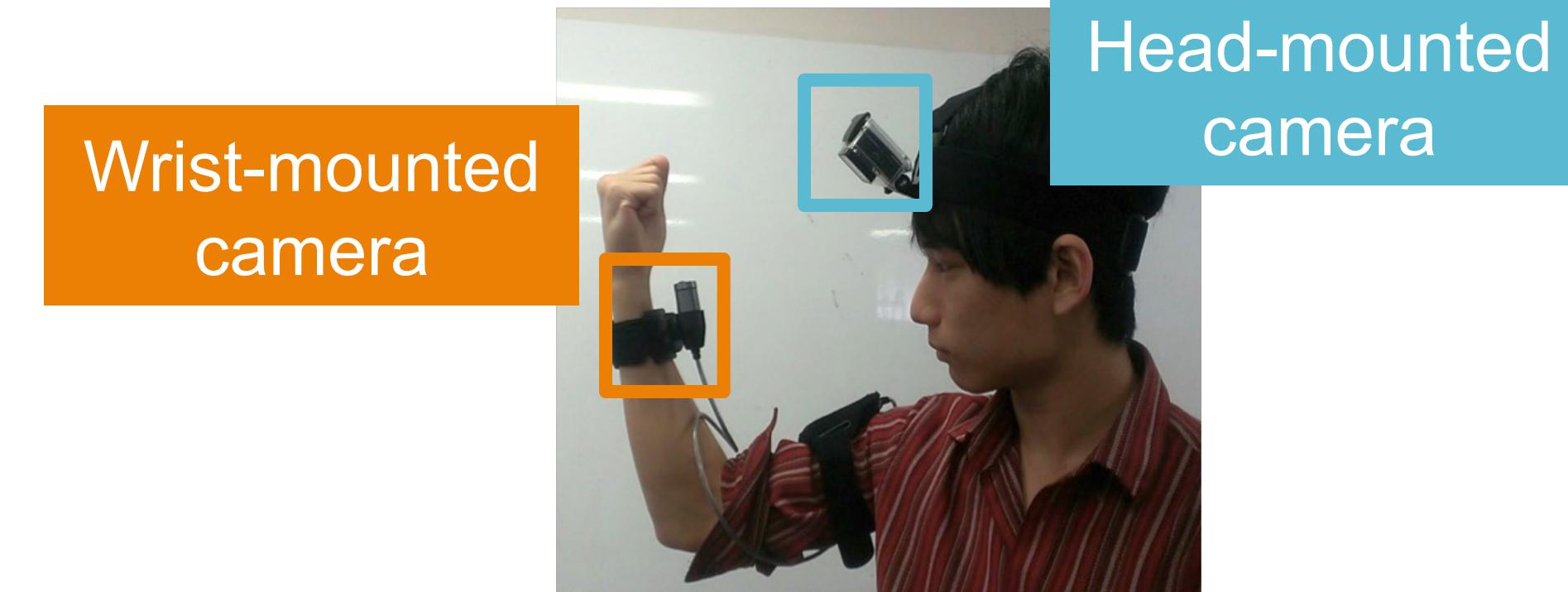
Discriminative Spatial Pyramid [T. Harada, et al., CVPR 2011]



Weight vector is obtained by PLS or Fisher

Dataset

- 20 subjects
- 20 different houses
- 23 daily activities
- 6.5 hours
- Publicly available!



Download:

<http://www.mi.t.u-tokyo.ac.jp/static/projects/milad/>

Experimental results

Video Features	WMC	HMC
LCD + VLAD [Z. Xu, et al., CVPR 2015]	78.6	62.4
DSAR (ours)	82.0	61.6
DSTAR (ours)	83.7	62.0

Table1: Mean classification accuracy of proposed method

Video Features	WMC	HMC
iDT + FV [H. Wang & C. Schmid, ICCV 2013]	73.6	78.1
LCD & iDT + FV	84.1	80.5
DSTAR & iDT + FV (ours)	85.5	80.2
DSTAR (WMC) & iDT + FV (HMC) (ours)		89.7

Table2: Mean classification accuracy of combining CNN-based descriptors with motion features

