Additional Material for

"Action Tube Extraction based 3D-CNN for RGB-D Action Recognition"

Zineng Xu¹ (<u>xuzineng7@126.com</u>), Verónica Vilaplana² (<u>veronica.vilaplana@upc.edu</u>), Josep Ramon Morros² (<u>ramon.morros@upc.edu</u>)

1. Per-class accuracies for the OA1/OA2 datasets:

OA1 – action class	Accuracy	OA2 – action class	Accuracy	
answering-phones	67.8%	asking-and-away	75.9 %	
arranging-files	84.7 %	called-away	87.9 %	
eating	96.6 %	carrying	87.9 %	
moving-objects	98.3 %	chatting	67.2 %	
going-to-work	96.7 %	delivering	63.4 %	
finding-objects	83.1 %	eating-and-chatting	98.3 %	
mopping	98.3 %	having-guest	91.4 %	
sleeping	98.3 %	seeking-help	87.9 %	
taking-water	100 %	shaking-hands	89.7 %	
wandering	93.2 %	showing	72.4 %	

2. Per-class accuracies for the NTU RGB-D dataset

NTU RGB-D action class	Accuracy
drink water	96.65%
eat meal/snack	80.91%
brushing teeth	87.61%
brushing hair	94.75%
drop	97.38%
pickup	98.46%
throw	92.86%
sitting down	98.55%
standing up (from sitting position)	98.10%
clapping	83.62%
reading	85.88%
writing	85.43%
tear up paper	97.10%
wear jacket	100%
take off jacket	95.93%

¹ Computer Science Departament, Universitat Politècnica de Catalunya (UPC)

² Signal Theory and Communications Department, <u>GPI</u> at <u>IDEAI</u> research group (UPC)

wear a shoe	00 250/
	88.25%
take off a shoe	75.49%
wear on glasses	96.38%
take off glasses	98. 55 %
put on a hat/cap	98.55 %
take off a hat/cap	97.46%
cheer up	96.29%
hand waving	94.48%
kicking something	96.01%
put something inside pocket	94.13%
hopping (one foot jumping)	96.38%
jump up	99.64%
make a phone call/answer phone	96.74%
playing with phone/tablet	94.48%
typing on a keyboard	96.65%
pointing to something with finger	94.84%
taking a selfie	95.29%
check time (from watch)	97.10%
rub two hands together	87.14%
nod head/bow	96.01%
shake head	100%
wipe face	88.23%
salute	95.20%
put the palms together	89.41%
cross hands in front (say stop)	96.20%
sneeze/cough	82.91%
staggering	98.19%
falling	96.29%
touch head (headache)	86.87%
touch chest (stomachache/heart pain)	96.29%
touch back (backache)	96.38%
touch neck (neckache)	94.84%
nausea or vomiting condition	86.16%
use a fan/feeling warm	85.17%
punching/slapping other person	95.48%
kicking other person	87.87%
pushing other person	96.65%
pat on back of other person	91.22%
point finger at the other person	96.65%
hugging other person	99.64%
giving something to other person	92.94%
touch other person's pocket	97.46%
handshaking	95.29%
walking towards each other	97.83%
walking apart from each other	96.74%

3. Confusion matrices for the OA1 dataset

answering-phones	0.678	0.017	0.068				0.034		0.136	0.068
arranging-files		0.847				0.119	0.017		0.017	
eating	0.034		0.966							
moving-objects				0.983						0.017
going-to-work				0.017	0.967					0.017
finding-objects		0.153				0.831			0.017	
mopping			0.017				0.983			
sleeping				0.017				0.983		
taking-water									1	
wandering				0.017					0.051	0.932
		an,	Caling	Thou;	Sojn.	Tingi	Nopping Obje	s leeping	(SACIO)	Wang
		arrangi	R. IIIes S	130	Soing, objects	finding.	Objects Oits		taking,	wandering

4. Confusion matrix for the OA2 dataset

asking-and-away	0.759	0.086			0.052		0.017	0.069	0.017	
called-away	0.052	0.879	0.034		0.034					
carrying		0.121	0.879							
chatting	0.034			0.672		0.207			0.069	0.017
delivering	0.155	0.069			0.634			0.138		
eating-and-chatting						0.983	0.017			
having-guest	0.017						0.914	0.069		
seeking-help	0.034	0.034			0.052			0.879		
shaking-hands				0.052	0.017				0.897	0.034
showing	0.017			0.034		0.086	0.017		0.121	0.724
		^	^	^		^	^	^	^	^

5. Confusion matrix for the NTU RGB-D dataset

