ROSE Rapid-Rich Object Search Lab
博 云 搜 索 实 验 室

- HOME
- DATASETS
  - Recaptured Images
  - Video Object Instance
  - Action Recognition
  - SIR^2 Benchmark
  - ROSE-Youtu Face Liveness Detection
  - NTU CCTV-Fights
  - Warwick-NTU Multi-camera Forecasting (WNMF) database

Action Recognition Datasets: "NTU RGB+D" Dataset and "NTU RGB+D 120" Dataset

This page introduces two datasets: "NTU RGB+D" and "NTU RGB+D 120".
"NTU RGB+D" contains 60 action classes and 56,880 video samples.
"NTU RGB+D 120" extends "NTU RGB+D" by adding another 60 classes and another 57,600 video samples, i.e., "NTU RGB+D 120" has 120 classes and 114,480 samples in total.
These two datasets both contain RGB videos, depth map sequences, 3D skeletal data, and infrared (IR) videos for each sample. Each dataset is captured by three Kinect V2 cameras concurrently.
The resolutions of RGB videos are 1920x1080, depth maps and IR videos are all in 512x424, and 3D skeletal data contains the 3D coordinates of 25 body joints at each frame.

1. Action Classes

The actions in these two datasets are in three major categories: daily actions, mutual actions, and medical conditions, as shown in the tables below.
Note: actions labelled from A1 to A60 are contained in "NTU RGB+D", and actions labelled from A1 to A120 are in "NTU RGB+D 120".

1.1 Daily Actions (82)

A1: drink water	A2: eat meal	A3: brush teeth	A4: brush hair
A5: drop	A6: pick up	A7: throw	A8: sit down
A9: stand up	A10: clapping	A11: reading	A12: writing
A13: tear up paper	A14: put on jacket	A15: take off jacket	A16: put on a shoe
A17: take off a shoe	A18: put on glasses	A19: take off glasses	A20: put on a hat/cap
A21: take off a hat/cap	A22: cheer up	A23: hand waving	A24: kicking something
A25: reach into pocket	A26: hopping	A27: jump up	A28: phone call
A29: play with phone/tablet	A30: type on a keyboard	A31: point to something	A32: taking a selfie
A33: check time (from watch)	A34: rub two hands	A35: nod head/bow	A36: shake head
A37: wipe face	A38: salute	A39: put palms together	A40: cross hands in front
A61: put on headphone	A62: take off headphone	A63: shoot at basket	A64: bounce ball
A65: tennis bat swing	A66: juggle table tennis ball	A67: hush	A68: flick hair
A69: thumb up	A70: thumb down	A71: make OK sign	A72: make victory sign
A73: staple book	A74: counting money	A75: cutting nails	A76: cutting paper
A77: snap fingers	A78: open bottle	A79: sniff/smell	A80: squat down
A81: toss a coin	A82: fold paper	A83: ball up paper	A84: play magic cube
A85: apply cream on face	A86: apply cream on hand	A87: put on bag	A88: take off bag
A89: put object into bag	A90: take object out of bag	A91: open a box	A92: move heavy objects
A93: shake fist	A94: throw up cap/hat	A95: capitulate	A96: cross arms
A97: arm circles	A98: arm swings	A99: run on the spot	A100: butt kicks
A101: cross toe touch	A102: side kick	-	-

1.2 Medical Conditions (12)

A41: sneeze/cough	A42: staggering	A43: falling down	A44: headache
A45: chest pain	A46: back pain	A47: neck pain	A48: nausea/vomiting
A49: fan self	A103: yawn	A104: stretch oneself	A105: blow nose

1.3 Mutual Actions / Two Person Interactions (26)

A50: punch/slap	A51: kicking	A52: pushing	A53: pat on back
A54: point finger	A55: hugging	A56: giving object	A57: touch pocket
A58: shaking hands	A59: walking towards	A60: walking apart	A106: hit with object
A107: wield knife	A108: knock over	A109: grab stuff	A110: shoot with gun
A111: step on foot	A112: high-five	A113: cheers and drink	A114: carry object

A115: take a photo	A116: follow	A117: whisper	A118: exchange things
A119: support somebody	A120: rock-paper-scissors	-	-

2. Size of Datasets

To ease the downloading, we separate the modalities of the samples into different files. The size of each modality is shown in the below table:

Data Modality	**"NTU RGB+D"**	**"NTU RGB+D 120"**
3D skeletons (body joints)	5.8 GB	5.8+4.5 GB
Masked depth maps*	83 GB	83+64 GB
Full depth maps	886 GB	886+549 GB
RGB videos	136 GB	136+124 GB
IR data	221 GB	221+168 GB
Total	**1.3 TB**	**2.3 TB**

*Masked depth maps are the foreground masked version of the depth maps. Masking is done based on the locations of the detected body joints, to remove the background and less important parts of the depth maps and to improve the compression rate.

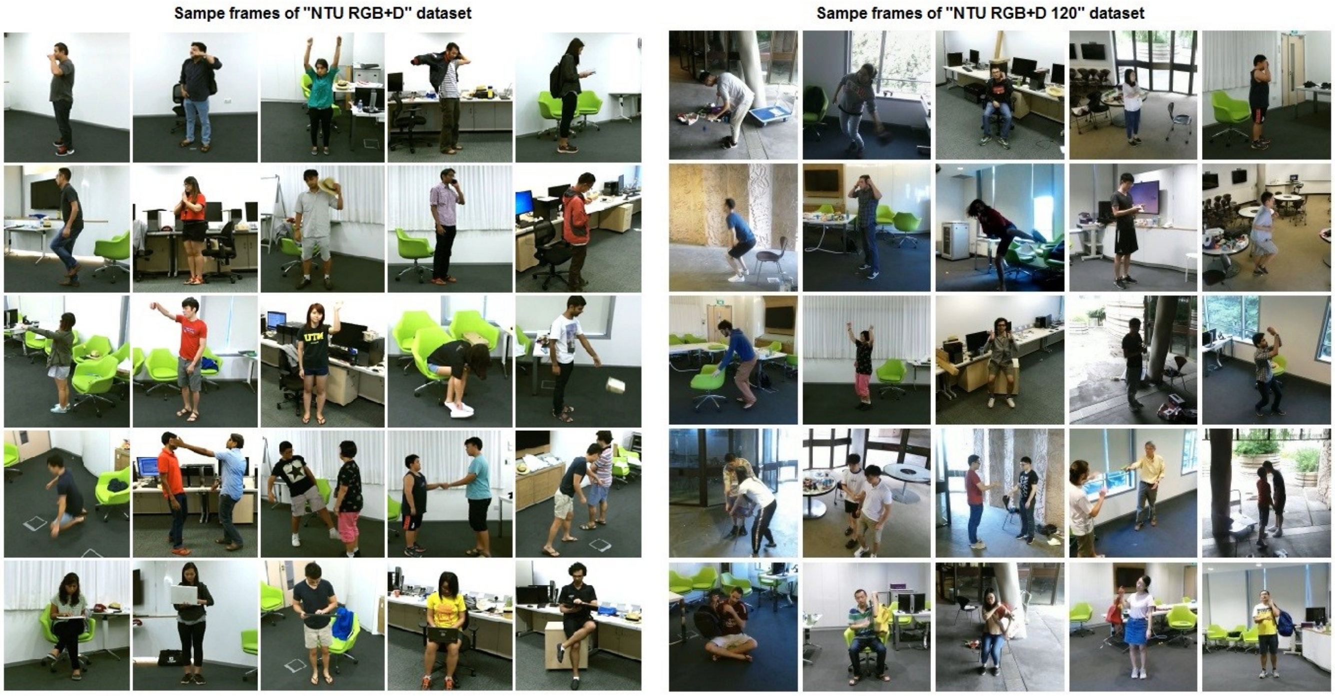
3. How to Download Datasets

Please click on the "Request Dataset" button at the bottom of this page. We will then send the LoginID to you for downloading the datasets.
The LoginID can be used for both "NTU RGB+D" and "NTU RGB+D 120".

4. More Information (FAQs and Sample Codes)

We provide more information about the data, answers to FAQs, samples codes to read the data, and the latest published results on our datasets here.

5. Sample Videos


Sampe frames of "NTU RGB+D" dataset
Sampe frames of "NTU RGB+D 120" dataset

6. Usage for Academic Reseach

The datasets are released for academic research only, and are free to researchers from educational or research institutes for non-commercial purposes.

If interested, please click on the "Request Dataset" button at the bottom of this page. We will then send the LoginID to you for downloading the datasets. The LoginID can be used for both "NTU RGB+D" and "NTU RGB+D 120".

7. Related Publications

All publications using "NTU RGB+D" or "NTU RGB+D 120" Action Recognition Database or any of the derived datasets(see Section 8) should include the following acknowledgement: "(Portions of) the research in this paper used the NTU RGB+D (or NTU RGB+D 120) Action Recognition Dataset made available by the ROSE Lab at the Nanyang Technological University, Singapore."

Furthermore, these publications should cite the following papers:

- **Amir Shahroudy, Jun Liu, Tian-Tsong Ng, Gang Wang, "NTU RGB+D: A Large Scale Dataset for 3D Human Activity Analysis", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016** [PDF] [bibtex].
- **Jun Liu, Amir Shahroudy, Mauricio Perez, Gang Wang, Ling-Yu Duan, Alex C. Kot, "NTU RGB+D 120: A Large-Scale Benchmark for 3D Human Activity Understanding", IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2019.** [PDF] [bibtex].

Some related works on RGB+D action recognition:

- Amir Shahroudy, Tian-Tsong Ng, Qingxiong Yang, Gang Wang, "Multimodal Multipart Learning for Action Recognition in Depth Videos", TPAMI, 2016.
- Amir Shahroudy, Tian-Tsong Ng, Yihong Gong, Gang Wang, "Deep Multimodal Feature Analysis for Action Recognition in RGB+D Videos" TPAMI, 2018.
- Amir Shahroudy, Gang Wang, Tian Tsong Ng, "Multi-modal Feature Fusion for Action Recognition in RGB-D Sequences", ISCCSP, 2014.
- Jun Liu, Amir Shahroudy, Dong Xu, Gang Wang, "Spatio-Temporal LSTM with Trust Gates for 3D Human Action Recognition", ECCV, 2016.
- Jun Liu, Gang Wang, Ping Hu, Ling-Yu Duan, Alex C. Kot, "Global Context-Aware Attention LSTM Networks for 3D Action Recognition", CVPR, 2017.
- Jun Liu, Amir Shahroudy, Dong Xu, Alex C. Kot, Gang Wang, "Skeleton-Based Action Recognition Using Spatio-Temporal LSTM Network with Trust Gates", TPAMI, 2018.
- Jun Liu, Gang Wang, Ling-Yu Duan, Kamila Abdiyeva, Alex C. Kot, "Skeleton-Based Human Action Recognition with Global Context-aware Attention LSTM Networks", TIP, 2018.
- Jun Liu, Amir Shahroudy, Gang Wang, Ling-Yu Duan, Alex C. Kot, "Skeleton-Based Online Action Prediction Using Scale Selection Network", TPAMI, 2019.

8. Derived Works Based on NTU RGB+D Dataset

Below are some datasets that are derived from NTU RGB+D dataset:

LSMB19: A Large-Scale Motion Benchmark for Searching and Annotating in Continuous Motion Data Streams (http://mocap.fi.muni.cz/LSMB).
J. Sedmidubsky, P. Elias, P. Zezula, "Benchmarking Search and Annotation in Continuous Human Skeleton Sequences", ICMR, 2019.

Note: Users of these derived works should also cite the papers found here (see Section 7 on Related Publications)

Request Dataset **Authorised to Download**