SUPPLEMENTARY DOCUMENT FOR HOLOGESTURE: A MULTIMODAL DATASET FOR HAND GESTURE RECOGNITION ROBUST TO HAND TEXTURES ON HEAD-MOUNTED MIXED-REALITY DEVICES

Jeongwoo Park   *Je Hyeong Hong*[⋆]

Department of Electronic Engineering, Hanyang University, Seoul 04763, South Korea

ABSTRACT

In this supplementary document, we describe the HoloGesture dataset and the evaluation framework. The dataset can be downloaded from `https://drive.google.com/file/d/1HQSGy8I8-efcsgI6EUu1L61CYpSjTRwM/view?usp=sharing`, and the code is available at `https://github.com/hellojpark/hologesture`.

1. DATASET

1.1. Intended uses

HoloGesture is a dataset acquired considering the use of commercially available MR devices in medical settings where surgical latex gloves are worn. To recognize hand gestures in MR devices while wearing surgical latex gloves, we utilized the commercially available MR device, HoloLens 2, to acquire the data. Unlike existing datasets, we have captured hand gesture data not only with bare hands but also while wearing two types of surgical latex gloves.

1.2. Acquisition

During the data acquisition process, we obtained consent from individuals performing the hand gestures, and ensured that the data was collected without infringing upon the rights of the individuals depicted.

The dataset comprises a total of 27 classes, with detailed actions illustrated in Fig. 1. Additionally we capture hand gesture including bare-handed and wearing two types of medical latex glove. Bare-handed and two types of gloves is depicted in Fig. 2.

1.3. Dataset construction

The code used for acquiring and preprocessing the Hololens 2 data was adapted from HoloLens2ForCV, which is licensed under 'MIT license'.

The video data representing hand gestures is stored frame by frame. The entire process of performing a single gesture is saved as one video, and the portions of the video corresponding to the hand gestures are marked with the starting and ending frame numbers in '.lst' extension files. The Depth, Depth$_{RGB}$, RGB$_{depth}$, and RGB data mentioned in the paper are all stored with a '.png' extension. The provided data is organized within folders named 'depth', 'rgb_based_depth', 'depth_based_rgb', and 'PV_aligned' respectively, where the videos are stored.

The data files are provided in the following order: "subject number → glove type → class number → trial number → data type". "subject1/blue/class22/r2/depth" is one of directory example of Hologesture folder.

The data is loaded using information specified in '.json' and '.lst' extension files. The '.json' file contains information about the type of data, the overarching folder path encompassing the data, and the path to the '.lst' extension file. The '.lst' file details the specific path to the video, label information, and frame names.

By utilizing the information documented in each file, the file path, label, and frame numbers corresponding to the hand gestures are stored to load the data. The process of loading the data is detailed in the 'read_data.py' and 'Hololens_data.py' files provided with each code. This code has been restructured based on the resources available at refer code

2. REPRODUCIBILITY FRAMEWORK

2.1. Model

We conducted experiments using C3D and ReT as baseline models. The C3D model was based on C3D, and the ReT model was reconstructed by referencing ReT. The experimental replication can be implemented by following the code available at Our Code and the instructions in the 'README.md'.

2.2. Evaluation

We separate model codes according to data modalities. Three types of network codes are provided for 'depth', 'RGB', and 'multi modal fusion'. As a result, we offer 6 network codes

⋆Corresponding author

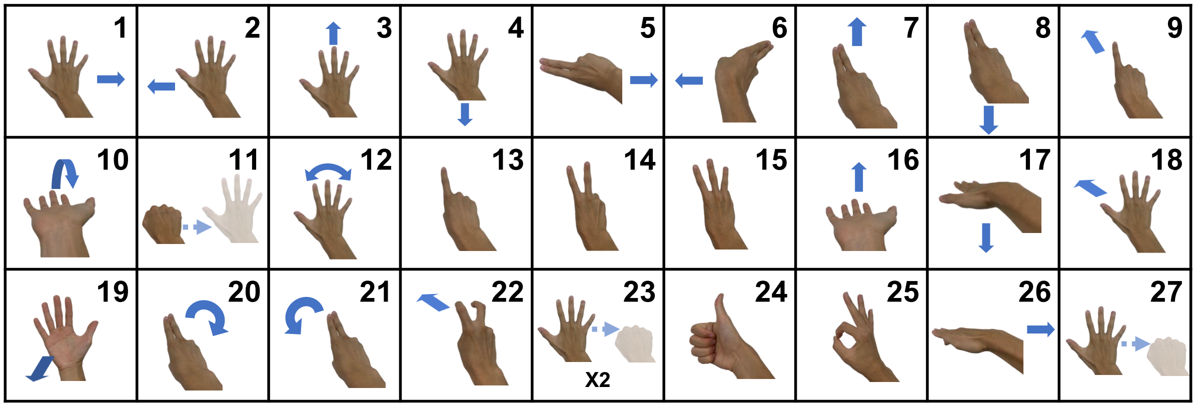

Fig. 1. The 27 classes of hand gesture in HoloGesture

(a) Bare hand (b) Blue glove (c) White glove

Fig. 2. Three kinds of RGB image when wearing two types of latex glove and bare-handed

within the 'model' folder on our GitHub URL. Our experimental results can be evaluated and reproduced using the provided pretrained model files. A link to download the pretrained models is available on our GitHub URL. Additionally, we include commands in the 'README.md' file to facilitate easy execution of our code. For C3D training, it is necessary to download another pretrained model named 'c3d-pretrained.pth', which should be placed in the 'model/c3d' folder for fine-tuning.

When attempting to evaluate our code, there are several parsers available: 1) '–hypes' requires the directory of a '.json' file. You should select the appropriate '.json' file based on the modality data you intend to use, ensuring to choose the correct 'specific_path' name within the '.json' file. For fusion networks, two types of specific paths exist: 'rgb_specific_path' and 'depth_specific_path', which pertain to the filenames of modality data videos. 2) '–resume' requires the directory of a pretrained file. 3) For utilizing a fusion network, the '–fusionkind' parser is necessary. It requires the name of the fusion approach, which includes 'late_fusion' and 'feature_fusion'. 4) In the case of the ReT network, the '–phase' parser is needed. It requires specifying one of two modes: train or test.

2.3. Training

For training the code, the '–resume' parser is not necessary. The types of parsers are detailed in 'main.py' for training ReT and 'train.py' for C3D. It is crucial to carefully review the '.json' file, which includes several significant variables for training: 'data_path', 'crop_data_path', 'specific_path', 'width', 'heigth', 'save_dir' and 'save_name'. 1) 'data_path' refers to the HoloGesture dataset directory. Upon downloading our dataset via the provided URL link, you will receive a folder named holodataset. You should enter the holodataset directory in 'data_path'. 2) 'crop_data_path' requires the directory of a '.lst' file. There are four types of '.lst' files: 'crop_lst' for utilizing the entire HoloGesture dataset. 'color_mixed_crop_lst' for the 'color_mixed' dataset. 'bare_crop_lst' for the 'bare_only' dataset. 'white_crop_lst' and 'blue_crop_lst' for datasets exclusively containing white or blue latex glove data, respectively. This is related to the test dataset. 3) 'specific_path' is about the folder name of modality data. There are five folder names: 'range', 'depth', 'depth_based_rgb', 'PV_aligned', and 'rgb_based_depth', each corresponding to a specific modality data: range, depth, RGB$_{depth}$, RGB, and depth$_{RGB}$. 4) The 'width' and 'height' values should be adjusted based on the modality data selected for training: For depth or RGB$_{depth}$ data, set width to 320 and height to 288. For RGB or depth$_{RGB}$, set width to 760 and height to 428. For range data, both width and height should be set to 512. 5) 'save_dir' specifies the directory where results are saved. 'save_name' is the name of the pretrained model file.

We detail the content on our GitHub URL. Additionally, we continuously manage and update changes in the 'README.md' file.

3. ADDITIONAL EXPERIMENTAL RESULTS

We additionally provide class-wise classification results using multimodal data. We split test dataset into 27 subsets, each containing 30 videos. Each subset includes only one class. We evaluate the baseline models, trained using the HoloGesture training dataset, with these 27 test subsets. The experimental results for each modality are presented in Tables 1, 2, 3, and 4.

Table 1. Class-wise classification accuracy.

Class	Depth		
	D'Eusanio *et al.*	Tran *et al.*	Zhou *et al.*
1	96.67%	83.33%	100.0%
2	86.67%	83.33%	100.0%
3	70.00%	90.00%	100.0%
4	93.33%	83.33%	100.0%
5	96.67%	70.00%	90.00%
6	96.67%	73.33%	100.0%
7	100.0%	76.67%	100.0%
8	90.0%	63.33%	90.00%
9	83.33%	56.67%	76.67%
10	96.67%	73.33%	96.67%
11	100.0%	76.67%	100.0%
12	83.33%	73.33%	100.0%
13	80.00%	46.67%	92.86%
14	100.0%	13.33%	96.67%
15	96.67%	40.00%	96.67%
16	100.0%	80.00%	93.33%
17	93.33%	83.33%	100.0%
18	100.0%	80.00%	96.67%
19	96.67%	96.67%	100.0%
20	96.67%	90.00%	96.67%
21	86.67%	83.33%	100.0%
22	90.00%	60.00%	96.67%
23	96.67%	86.67%	100.0%
24	100.0%	100.0%	100.0%
25	93.33%	76.67%	100.0%
26	96.67%	36.67%	100.0%
27	100.0%	73.33%	100.0%

Table 2. Class-wise classification accuracy.

Class	RGB		
	D'Eusanio *et al.*	Tran *et al.*	Zhou *et al.*
1	100.0%	83.33%	100.0%
2	100.0%	86.67%	100.0%
3	83.33%	33.33%	100.0%
4	86.67%	13.33%	100.0%
5	93.33%	26.67%	100.0%
6	96.67%	83.33%	100.0%
7	96.67%	80.00%	100.0%
8	96.67%	83.33%	100.0%
9	66.67%	46.67%	76.67%
10	96.67%	86.67%	100.0%
11	83.33%	53.33%	100.0%
12	96.67%	46.67%	100.0%
13	100.0%	13.33%	82.14%
14	100.0%	10.00%	96.67%
15	100.0%	66.67%	86.67%
16	83.33%	56.67%	96.67%
17	60.00%	70.00%	100.0%
18	80.00%	76.67%	96.67%
19	100.0%	76.67%	100.0%
20	86.67%	66.67%	100.0%
21	96.67%	76.67%	100.0%
22	86.67%	40.00%	100.0%
23	100.0%	80.00%	100.0%
24	100.0%	73.33%	100.0%
25	96.67%	83.33%	100.0%
26	93.33%	60.00%	100.0%
27	86.67%	80.00%	100.0%

Table 3. Class-wise classification accuracy.

Class	$\text{Depth}_{\text{RGB}}$		
	D'Eusanio *et al.*	Tran *et al.*	Zhou *et al.*
1	100.0%	100.0%	96.67%
2	100.0%	100.0%	100.0%
3	86.67%	36.67%	76.67%
4	76.67%	66.67%	96.67%
5	90.00%	53.33%	66.67%
6	90.00%	86.67%	96.67%
7	93.33%	86.67%	96.67%
8	73.33%	73.33%	93.33%
9	73.33%	43.33%	46.67%
10	96.67%	56.67%	86.67%
11	60.00%	50.00%	93.33%
12	66.67%	36.67%	93.33%
13	86.67%	53.33%	92.86%
14	96.67%	76.67%	90.00%
15	93.33%	23.33%	93.33%
16	100.0%	40.00%	96.67%
17	66.67%	63.33%	96.67%
18	90.00%	50.00%	86.67%
19	100.0%	93.33%	96.67%
20	96.67%	70.00%	100.0%
21	93.33%	86.67%	100.0%
22	96.67%	73.33%	96.67%
23	96.67%	96.67%	90.00%
24	100.0%	66.67%	66.67%
25	100.0%	60.00%	70.00%
26	93.33%	50.00%	75.00%
27	96.67%	63.33%	73.33%

Table 4. Class-wise classification accuracy.

Class	$\text{RGB}_{\text{Depth}}$		
	D'Eusanio *et al.*	Tran *et al.*	Zhou *et al.*
1	100.0%	83.33%	100.0%
2	100.0%	90.00%	100.0%
3	73.33%	63.33%	100.0%
4	86.67%	63.33%	96.67%
5	96.67%	16.67%	93.33%
6	93.33%	86.67%	100.0%
7	66.67%	86.67%	90.00%
8	86.67%	86.67%	90.00%
9	56.67%	30.00%	56.67%
10	100.0%	70.00%	100.0%
11	100.0%	46.67%	100.0%
12	90.00%	60.00%	100.0%
13	100.0%	46.67%	78.57%
14	86.67%	26.67%	100.0%
15	100.0%	66.67%	93.33%
16	100.0%	53.33%	96.67%
17	93.33%	63.33%	100.0%
18	83.33%	66.67%	100.0%
19	100.0%	83.33%	96.67%
20	96.67%	86.67%	100.0%
21	93.33%	86.67%	96.67%
22	96.67%	66.67%	96.67%
23	93.33%	80.00%	100.0%
24	100.0%	70.00%	100.0%
25	100.0%	80.00%	100.0%
26	90.00%	73.33%	95.00%
27	100.0%	70.00%	100.0%