3D Daily Activity Recognition Dataset for Elderly-care Robots

Dohyung Kim, Jinhyeok Jang, Minsu Jang, Jaeyeon Lee, Jaehong Kim

Abstract— **In the era of deep learning, the lack of large-scale datasets suitable for robots has been a major obstacle to advanced robotic intelligence researches. To boost action recognition studies for elder care robots, we propose a new dataset with three major characteristics: 1) practical action categories based on the analysis of the real daily lives of the elderly; 2) realistic data collection which reflects the robot's working environment and the diverse service situations; and 3) a large-scale dataset which overcomes the limits of the current 3D activity analysis benchmark datasets. The presented RGB-D dataset contains 122,620 samples including RGB videos, depth maps, and skeleton sequences. In variant collection settings, 100 subjects perform 55 daily activities. We hope that the proposed dataset, which comprehensively considers the elderly, the robots and the environment in which they interact, can contribute to the advancement of robot intelligence.**

I. Introduction

Despite the large number of publicly available datasets, there is a great lack of adequate data for robots to recognize daily activities of human users. Most datasets have no consideration for the robotic environment in which humans and robots live together. Furthermore, there is no large-scale visual dataset at all that deals with the everyday behavior of the elderly. The absence of datasets centered on robots and humans has been a serious impediment to robot intelligence researches, especially for elder care robots. To solve the shortage of datasets, we present a large-scale new RGB-D dataset for robots to recognize daily activity of the elderly: ETRI-Activity3D. Unique characteristics and advantages of the proposed dataset over the existing ones are as follows.

A new visual dataset based on observations of the daily activities of the elderly: A closer understanding of what older people actually do in their daily lives is important for determining practical action categories. We visit the homes of 53 elderly people over the age of 70 and carefully monitor and document their daily behavior from morning to night. Based on the most frequent behaviors observed, 55 action classes are defined.

A realistic dataset considering the service situation of human care robots: The aim of the presented dataset is to be utilized in practical research that can be applied to real-world environments. Therefore, in designing the dataset, possible situations that can occur when robots are in service are taken into account. First of all, the data acquisition environment simulating the living conditions of the elderly is set up in an apartment, not in a lab. We acquire data from possible locations where robots can be present when serving humans.

D. Kim, J. Jang, M. Jang, J. Lee and J. Kim are with Human Robot Interaction Research Lab, Electronics and Telecommunications Research Institute, Daejeon, Republic of Korea ({dhkim008, jjh6297, minsu, leejy, jhkim504}@etri.re.kr).

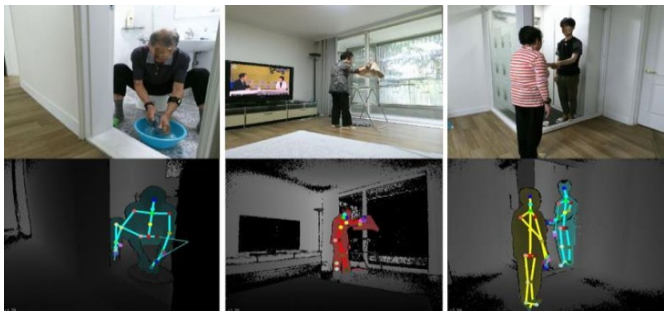

Figure 1. Sample frames are displayed together with the corresponding depth map and skeleton information. Actions (from left to right): *washing a towel by hands, hanging out laundry, hand shaking*. The full dataset is available for download at: https://ai4robot.github.io/etri-activity3d-en.

For example, in small spaces such as kitchens and bathrooms, our dataset can include scenes where humans show their backs while performing actions. It is not easy to recognize the behavior from the back, but this is a realistic situation that can occur frequently. These considerations centered on the robotic environment make the presented dataset more challenging and practical.

A large-scale RGB-D action recognition dataset that overcomes the limitations of previous datasets: Building 3D datasets is not easy work because there is no proper way to leverage video sharing services or crowdsourcing as in the case of 2D video datasets. These difficulties have limited previous 3D datasets in several aspects. The lack of subjects and action categories and monotonous environmental conditions are major limitations that prevent precise comparative evaluation of various approaches. As reported in Table 1, the presented dataset overcomes all these limitations and consists of 112,620 samples, which is comparable to the current largest 3D action recognition dataset, NTU RGB+D 120 [6]. The large-scale dataset reflecting realistic variations enables extensive research from a variety of perspectives.

II. The ETRI-Activity3D Dataset

The dataset is collected by Kinect v2 sensors and consists of three synchronized data modalities: RGB videos, depth maps, and skeleton sequences. The resolution of RGB videos is 1920 × 1080. Depth maps are stored frame by frame in 512 × 424. Skeleton information contains locations of 25 body joints in the 3D space for tracked human bodies.

There are a total of 55 action categories in the dataset. Based on observations of older people's real lives, 52 daily actions are selected: eating, cleaning, reading, etc. Among them are five mutual actions: handshaking, hugging each other, etc. We also define three useful actions especially for human-robot interaction: waving, beckoning, and pointing.

TABLE I. COMPARISON BETWEEN ETRI-ACTIVITY3D AND OTHER PUBLICLY AVAILABLE DATASETS FOR 3D DAILY ACTIVITY RECOGNITION. DATA MODALITIES: (RGB)VIDEO, (D)EPTH, (S)KELETON, (I)R.

Datasets	#Samples	#Sub	#Act	Modalities
RGBD-HuDaAct [1]	1,189	30	13	RGBD
Act4^2 [2]	6,844	24	14	RGBD
CAD-120 [3]	120	4	10+10	RGBDS
Office Activity [4]	1,180	10	20	RGBD
UWA3D Multiview II [5]	1,075	10	30	RGBDS
NTU RGB+D 120 [6]	114,480	106	120	RGBDSI
Toyota Smarthome [7]	16,129	18	31	RGBDS
ETRI-Activity3D	**112,620**	**100**	**55**	RGBDS

A total of 100 subjects, comprising 50 senior citizens and 50 young people, participate in the data collection. The elderly subjects range in age from 64 to 88 and the average age is 77. The young subjects are all in their 20s and are 23 years old on average. There are 15 men and 35 women in the elderly subjects, and the number of men and women in the younger subjects is the same.

Considering the height of home robots, the shooting device is prepared with two Kinect sensors at heights of 70cm and 120cm. The four shooting devices are grouped together, and eight synchronized sensors in the group capture the subjects' action at the same time. Instead of placing the devices at fixed horizontal angular intervals, we place them in a position where the robot can appear inside the house. For actions that can be done anywhere (e.g., taking medicine and talking on the phone), we shoot them up to five times, changing the places where they might occur. In this way, we can provide further intra-class variation by containing different views and background conditions.

Here we summarize the main characteristics of ETRI-Activity3D: (1) The daily activities of the elderly were recorded from the robot's point of view. (2) Action classes were determined based on the observations of the daily activities of the elderly. (3) The dataset acquisition was performed in an apartment reflecting the living conditions of the elderly, not in a laboratory. (4) The probable service situation of the human-care robots was considered. (5) The dataset provides the large amount variation in views, distances and backgrounds. (6) The dataset is the second largest dataset in daily living activity recognition domain in terms of number of subjects and video samples.

III. EXTENSION OF THE DATASET

We are currently building an additional 3D daily activity dataset of the elderly in the home environment where the elderly actually live and also in the virtual environment simulating it. We are visiting homes of 30 elderly people and acquiring another 3D activity dataset from the robot's point of view. This dataset will be more challenging and more representative of the real-world. Virtual synthetic action dataset of the elderly is also generated using computer graphics techniques.

The aim of the proposed dataset is to be utilized in practical researches that can be applied to real world environments.

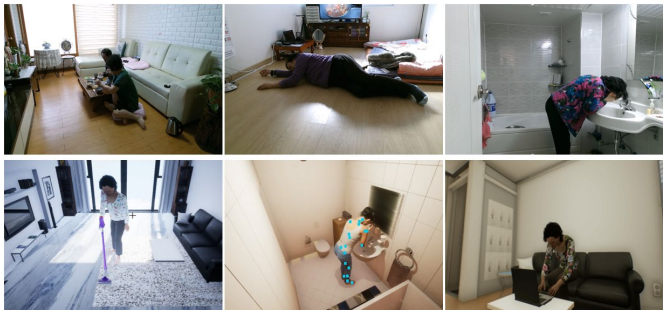

Figure 2. Sample frames of the additional datasets are displayed. First and second rows show examples from the actual home environment of the elderly and examples from the virtual environment, respectively.

These datasets of acquired in different environments enable a variety of domain adaptation studies. This additional dataset will be released in the near future to be utilized.

IV. CONCLUSION

The ETRI-Activity3D, a new large-scale 3D action recognition dataset, contains 112,620 video samples collected from 55 action classes performed by 100 subjects (50 elderlies and 50 adults). The dataset is expected to be useful for research on action recognition, robotic intelligence and elderly-care.

ACKNOWLEDGMENT

This work was supported by the ICT R&D program of MSIP/IITP. [2017-0-00162, Development of Human-care Robot Technology for Aging Society]. The protocol and consent of data collection were approved by the Institutional Review Board(IRB) at Suwon Science College, our joint research institute.

REFERENCES

[1] B. Ni, G. Wang, and P. Moulin, "Rgbd-hudaact: A color-depth video database for human daily activity recognition," in *IEEE Int. Conf. Computer Vision Workshops*, 2011.

[2] Z. Cheng, L. Qin, Y. Ye, Q. Huang, and Q. Tian, "Human daily action analysis with multi-view and color-depth data," in *European Conf. Computer Vision*, 2012.

[3] H. S. Koppula, R. Gupta, and A. Saxena, "Learning human activities and object affordances from rgb-d videos," in *Int. J. Robotics Research*, vol. 32, no. 8, 2013.

[4] L. Wang, Y. Qiao, and X. Tang, "Action recognition and detection by combining motion and appearance features," THUMOS, 2014.

[5] H. Rahmani, A. Mahmood, D. Huynh, and A. Mian, "Histogram of oriented principal components for cross-view action recognition," in *IEEE Tran. Pattern Analysis and Machine Intelligence*, vol. 32, no. 12, 2016.

[6] J. Liu, A. Shahroudy, M. L. Perez, G. Wang, L. Y. Duan, and A. K. Chichung, "Ntu rgb+ d 120: A large-scale benchmark for 3d human activity understanding," in *IEEE Tran. Pattern Analysis and Machine Intelligence*, 2019.

[7] S. Das, R. Dai, M. Koperski, L. Minciullo, L. Garattoni, F. Bremond, and G. Francesca, "Toyota smarthome: Real-world activities of daily living," in *IEEE Int. Conf. Computer Vision*, 2019.