

Human Gesture Recognition Performance Evaluation for Service Robots

Mi-Young Cho, Young-Sook Jeong

ETRI, 218 Gajeongno, Yuseong-gu, Daejeon, 34129, Korea

mycho@etri.re.kr, ysjeong@etri.re.kr

Abstract—Intelligent service robots use gesture recognition technology that utilizes the MS Kinect sensor to facilitate natural interactions between humans and robots. To evaluate gesture recognition performance in a real-life environment, we constructed a new gesture database that takes into account cluttered backgrounds, various distances and poses, and movement of robots, and then we evaluated the gesture recognition performance of commercial robots. In this paper, we seek to help consumers, robot manufacturers, and gesture recognition engine developers provide comparable results for the gesture recognition capabilities of service robots.

Keywords—Gesture recognition, service robot, Kinect sensor, performance evaluation, gesture dataset.

I. INTRODUCTION

A gesture is a form of non-verbal or non-vocal communication in which visible bodily actions communicate specific messages. With the development of new computer technology, it is now possible to implement various advanced interfaces. Advancements in keyboard and mouse technologies have made it possible to use gesture-based images to interact with robots.

Among many Human Robot Interaction (HRI) technologies, vision-based gesture recognition is a useful way for humans to interact with robots. It can be used efficiently in noisy environments, and it can provide active services by recognizing a user's intentions [6].

Gesture recognition technology for HRI involves both hand and body recognition. Technology using hand-gesture recognition has been mainly applied in the areas of gaming and sign language. Body-gesture recognition is used mainly in surveillance systems and healthcare.

Human gesture recognition from video sequences has been heavily studied because of its importance to applications that enhance patient monitoring systems for fall motion detection. It has also been studied for its usefulness in surveillance systems, motion analysis in sports, and for human behavior analysis.

Despite the amount of previous work, gesture recognition methods are still far from satisfactory for real-life applications. The quality of the captured images is sensitive to lighting conditions and cluttered backgrounds. Thanks to the recent development of inexpensive depth cameras, e.g., the Kinect

sensor, new opportunities for the study of gesture recognition have emerged.

The RoboCup@Home [8] initiative is a competition used to facilitate scientific exchange and to encourage efficient benchmarking. It uses a test arena that is configured as a realistic representation of a home environment. This competition includes several tests for specific abilities, and the successful integration of those abilities. Many different technologies, such as navigation, localization, object recognition and manipulation, and speech and gesture recognition are tested simultaneously in an integrated manner. Gesture recognition technology must demonstrate the ability to recognize and interpret human gestures.

The RoboCup@Home test is designed to simulate a real-life environment, but real-life environments are neither static nor standardized. There are limitations to the usefulness of the performance measurements unless the methodologies change over time. Current benchmarking tests evaluate gesture recognition separately from other features. Gesture recognition algorithms require the use of gesture datasets. Unlike standard, human gesture recognition technology, service robots are evaluated in various poses because they do not use fixed cameras; this ensures that performance evaluations are made that reflect the movement of the robot.

We considered the problem of accurately evaluating the gesture recognition performance of service robots in real service environments. In this paper, we present a new gesture database and testing method for the evaluation of service robots.

The remainder of this paper is organized as follows: Section 2 introduces the test method and the gesture database. Section 3 describes the construction of the gesture dataset, taking into account background, distance, pose, and movement. Section 4 presents and analyses the gesture recognition results obtained by using a commercial service robot. Section 5 concludes this paper.

II. RELATED WORKS

Many gesture recognition databases are used in evaluating human, gesture recognition technology. The Cornell University Robot Learning Lab [1, 2] constructed activity datasets that are used when capturing gestures using the Microsoft Kinect sensor. The datasets are comprised of RGB-D video sequences of humans performing activities. The

database is divided into CAD-60 and CAD-120. The CAD-60 dataset consists of 12 activities in five different environments (office, kitchen, bedroom, bathroom, and living room). Another dataset, CAD-120, contains long, daily activities subdivided into 10 high-level activities and 10 sub-activities. Even though this dataset was captured, despite various cluttered backgrounds, it could not reflect gesture variations due to a lack of test subjects.

The Microsoft Research Cambridge-12 Kinect Gesture Dataset [3] consists of sequences of human movements represented as body-part locations, with the associated gestures to be recognized by the system. They provide a useful dataset for both evaluating gesture recognition systems, and as a database for Kinect skeletal-track recordings and variations. Backgrounds are particularly straightforward for evaluating performance, because there is a performance gap between the test and real-life environments. Figure 1 shows an example of an MSRC-12 gesture dataset.



Figure 1. MSRC-12 gesture dataset.

The Advanced Digital Sciences Center (ADSC) has constructed an RGBD activities benchmark dataset by using the Microsoft Kinect sensor [4]. They collected 12 categories of human daily-activity data in a lab environment. Healthcare services provided the task definitions including, “make a phone call,” “mop the floor,” and so on. Backgrounds reflect an authentic service environment. The tests did not evaluate autonomous service robots because fixed cameras were used.

CharLearn [5] has many gesture video clips of the challenge. Figure 2 shows a sample RGB image and a grayscale rendering of depth images.

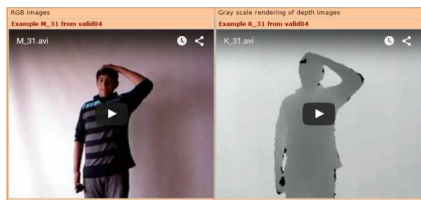


Figure 2. CharLearn gesture data.

There were 50,000 gestures recorded with the Kinect camera, including both RGB and depth videos, with image sizes of 240 ×320 pixels at 10 frames per second. They were recorded by 20 different users, and grouped in 500 batches of 100 gestures per batch. Each batch included 47 sequences of one to five gestures drawn from over 30 different small-gesture vocabularies. The selected small-gesture vocabularies each contained eight to fifteen gestures.

Existing gesture datasets were captured using simple unicolor backgrounds, and a front-centered posture; however, intelligent service robots in real-life environments operate in the background, and are complex, moving objects. In addition,

the robot should perform a function in the state, as well as the possible autonomous front gesture recognition is performed at a variety of angles moving. Upon considering the operating environments of robots, the necessity of building a new gesture database becomes clear.

III. GESTURE DATABASE

Current gesture databases are unsuitable for performance evaluations of service robots. We propose constructing a new gesture database that takes into account the combined factors of background, distance, pose, and movement in real-life environments.

A. Data Acquisition

This section describes in more detail how we acquired our data. We built our gesture database using the Kinect sensor (Ver. 2) for recognition of commonly used gestures. The color image resolutions were 1920×1080 pixels (full HD), and the resolutions of depth images were 640×480 pixels. Each depth pixel was a 16-bit integer. Both sequences were synchronized and the frame rates were 30 frames per second (FPS). The sequences also provided skeletal information for 25 joints and states.

We captured gesture video clips from 20 subjects; each video clip consisted of the same gestures repeated at least three times. We asked each subject to start in the position of attention. After performing each gesture, the subject was asked to return to the attention posture. We generated ground truth (start and end) based on the attention posture. Figure 3 shows the sample color and depth images.

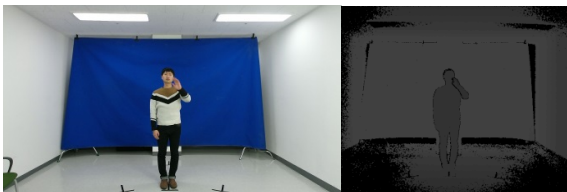


Figure 3. Kinect Color and depth images.

B. Variations

1) Background

We used three backgrounds to evaluate gesture recognition for service robots. The “studio” environment consisted of a unicolor background and no moving subjects. The “office” environment had general fluorescent lighting and no moving subjects, but it did have a cluttered background. To simulate a public space, we constructed a “lobby” with a background that contained both static and moving objects. Table 1 shows the sample images according to background.

TABLE 1. BACKGROUNDS

		
studio	office	lobby

2) Distance

A mobile robot capable of autonomous activity was employed, rather than a fixed camera, and it executed human-gesture recognition at various distances. Table 2 shows sample images according to distance (short distance is 1.5 m and long distance is 3.5 m).

TABLE 2. SAMPLE GESTURE IMAGES BASED ON DISTANCE



3) Pose

Service robots should be able to recognize gestures from various directions, as well as from the front. Current gesture databases cannot evaluate performance for the various poses that occur in a real environment. In our experiments, we took left and right-pose images as well as from the front. Table 3 shows the sample gesture images according to pose.

TABLE 3. SAMPLE GESTURE IMAGES BASED ON POSE



4) Movement

With new developments in service-robot technology, there has also been a move towards more widespread application of autonomous navigation in the security industry. Unlike in traditional human gesture recognition technology, gesture data collected from a mobile robot is needed in order to evaluate gesture recognition performance of robots in authentic service environments. Figure 4 shows the device and the environment we used to capture gestures from a moving robot. The device, which is on a rail, tracks both forward and backward movements at a constant speed.

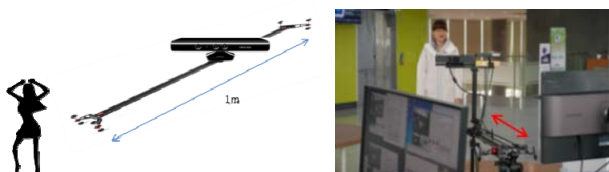


Figure 4. Mobile apparatus used for capturing gesture data.

In our tests, the movement speed was considered the normal speed of the robot with the subject gesturing while the rail was moving. Figure 5 shows video clips of background-based gestures captured with a forward-moving apparatus.

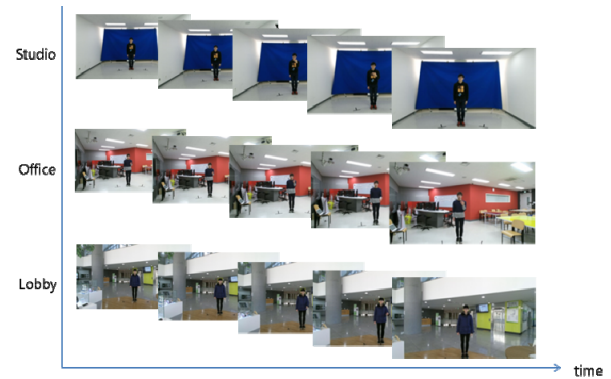


Figure 5. Sample gesture data based on movement.

IV. PERFORMANCE EVALUATION

A. Overview

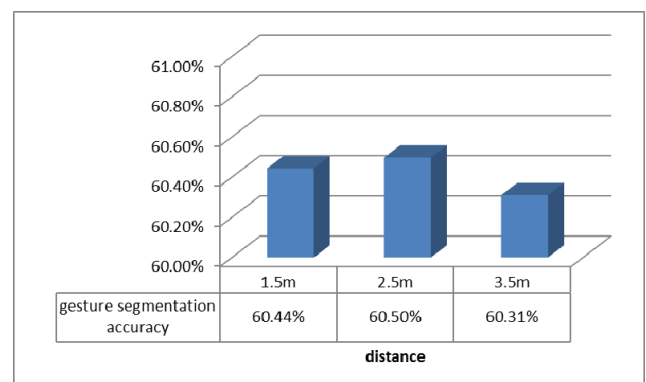
We tested commercial robots used for guiding visitors in a lobby. The test robot used a Kinect sensor to recognize a hand-greeting gesture.

The problem of human gesture recognition encompasses two areas: segmentation and recognition. Segmentation involves finding the starting and ending points of a legitimate gesture. Recognition involves matching the segmented gestures against a ground truth to determine which class it belongs to [7]. In our study, we focused on segmentation, specifically, how to robustly detect the gesture and then measure accuracy.

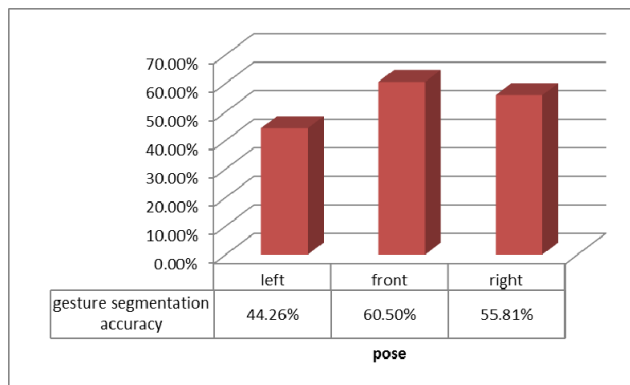
The test algorithm recognized human gestures and used skeletal sequence information. In order to evaluate the performance of a commercial service robot, we conducted an experiment using a hand-greeting gesture where we employed background, distance, pose, and movement to four test cases.

B. Results

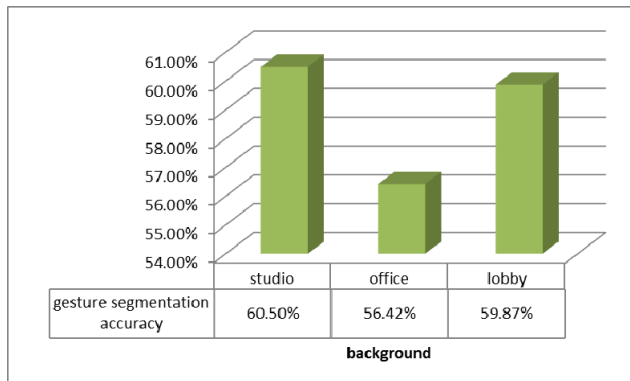
The test results can be seen in Fig. 6. In all four charts, the gesture segmentation accuracy is plotted along the y-axis, and the distance, pose, background, and direction of movement are plotted along the x-axis.



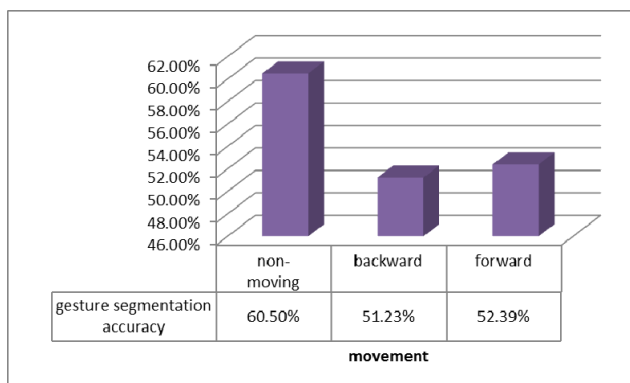
(a) distance



(b) pose



(c) background



(d) movement

Figure 6. Gesture recognition performance comparisons.

The best performance results were obtained at 2.5 m, and the lowest performance results were at 3.5 m. Gesture recognition from the front was superior to that of other directions in the pose test. Generally, gesture recognition performance deteriorated under moving conditions. The performance results were similar in the tests based on distance; however, other factors affected performance. In other words, gesture recognition technology in natural environments will need performance compensation.

V. CONCLUSIONS

In this paper, we dealt with the performance evaluation of human gesture recognition technology that is widely used for natural interaction between a human and a service robot. Unlike previous gesture databases, we constructed a gesture database that takes into account the authentic service environment. We then evaluated the gesture recognition performance of a commercial robot. Gesture recognition technology is capable of providing for the improvement of interactions between humans and robots. Seamless interactions between a user and a robot can improve reliability. Our test results indicate that our database can help consumers, robot manufacturers, and gesture recognition engine developers provide comparable results for the gesture recognition capabilities of service robots.

ACKNOWLEDGMENT

This work is supported in part by the R&D program of the Korea Ministry of Trade, Industry and Energy (MOTIE) and the Korea Evaluation Institute of Industrial Technology (KEIT) (Project: Technology development of service robot's performance and standardization for movement/manipulation/HRI/Networking, 10041834).

REFERENCES

- [1] SUNG, Jaeyong, et al. Unstructured human activity detection from rgbd images. In: Robotics and Automation (ICRA), 2012 IEEE International Conference on. IEEE, 2012. p. 842-849.
- [2] Koppula, Hema Swetha, Rudhir Gupta, and Ashutosh Saxena. "Learning human activities and object affordances from rgb-d videos." The International Journal of Robotics Research 32.8 (2013): 951-970.
- [3] FOTHERGILL, Simon, et al. Instructing people for training gestural interactive systems. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2012. p. 1737-1746.
- [4] NI, Bingbing; WANG, Gang; MOULIN, Pierre. Rgb-d-hudaact: A color-depth video database for human daily activity recognition. In: Consumer Depth Cameras for Computer Vision. Springer London, 2013. p. 193-208.
- [5] BARÓ, Xavier, et al. Chalearn looking at people 2015 challenges: Action spotting and cultural event recognition. In: Computer Vision and Pattern Recognition Workshops (CVPRW), 2015 IEEE Conference on. IEEE, 2015. p. 1-9.
- [6] LEE, Seong-Whan. Automatic gesture recognition for intelligent human-robot interaction. In: Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on. IEEE, 2006. p. 645-650.
- [7] YANG, Hee-Deok, et al. Gesture spotting and recognition for human-robot interaction. Robotics, IEEE Transactions on, 2007, 23.2: 256-270.
- [8] WISSEINTNER, Thomas, et al. RoboCup@ Home: Scientific competition and benchmarking for domestic service robots. Interaction Studies, 2009, 10.3: 392-426.
- [9] PATSADU, Orasa; NUKOOLKIT, Chakarida; WATANAPA, Bunthit. Human gesture recognition using Kinect camera. In: Computer Science and Software Engineering (JCSSE), 2012 International Joint Conference on. IEEE, 2012. p. 28-32.



Mi-Young Cho is a senior researcher at the Electronics and Telecommunications Research Institute. She received her B.S. degree in information and telecommunication engineering from Chosun University in 2002, and her MS and Ph.D. degrees in computer science from Chosun University in 2004 and 2008, respectively. Since 2009, she has been researching the testing and evaluation of service robots at the Electronics and Telecommunications Research Institute.



Young-Sook Jeong is a principal researcher at the Electronics and Telecommunications Research Institute. She received her B.S. degree in computer science from the Ewha Womans University in 1988, and her MS degree in electronic engineering from the Chungnam National University in 2001. She has research experience in the field of performance evaluation and standardization for service robots.