Assessment of Drug Picking Activity using RGB-D Camera

Yuta Ono, Oky Dicky Ardiansyah Prima
Graduate School of Software and Information Science, Iwate Prefectural University
152-52 Sugo, Takizawa, Iwate, Japan
e-mail: g236s001@s.iwate-pu.ac.jp, prima@iwate-pu.ac.jp

***Abstract*—Non-pharmacists have been allowed to pick drugs under the responsibility of pharmacists in order to reduce the burden on them in Japan. However, the activity tends to occur human errors since the name or shape of drugs are similar. While Bar-Code Medication Administration (BCMA) system and Automated Dispensing System (ADS) have been proposed to prevent such errors, these systems are cumbersome and costly. With the progress on human pose estimation technique using machine vision-based approach, it has become possible to measure the displacement and the posture of human body in real space. This approach makes us easy to measure the location of humans and other objects simultaneously. This study attempts to construct a drug picking activity judgement framework using RGB-D camera to detect the error easily at low-cost. This framework uses RGB-D camera to measure the location of hand landmarks and judges the activity by proposed judgement algorithm based on the position of those landmarks. In order to measure both hands accurately, we used Azure Kinect Body Tracking SDK and MediaPipe. Our experiments show that proposed framework is capable of the activity judgement on drug picking.**

Keywords-Medication administration error; 3D human pose estimation; MediaPipe; RGB-D camera; Azure Kinect.

I. INTRODUCTION

In Japan, non-pharmacists have become possible to perform the picking of drugs, such as Press Through Package (PTP) sheets, under the responsibility of pharmacists [1]. This is expected pharmacists to concentrate on more specialized tasks, such as checking prescriptions and providing medication guidance for patients. However, drug picking activities tend to occur human errors due to the similarity of drug names, shapes and so on [2]. The error during the activity may cause serious harm to patients and make the operator place a heavy burden. In addition, pharmacists need to check the drugs collected by operator. Therefore, there is a need for a method to prevent human errors during the activity.

Previous works have been proposed various method to prevent human errors during the activity. Bar-Code Medication Administration (BCMA) is a system that use bar-codes to identify such drugs, prescriptions, operators and verify that work is being performed correctly. This system is effective to prevent medication administration errors and it has been shown that introducing of the BCMA system can significantly reduce the error rate [3][4]. However, those methods are cumbersome because operator need to scan bar-code each time to check picking operation. Similar to the bar-code, Radio Frequency Identification (RFID) technology has been used for object identification. This technology uses a RFID reader to identify object with RFID tag. The reader can scan multiple tags simultaneously. However, as with BCMA system, we need to scan them in order to find human error during the activity. Automated Dispensing System (ADS) is a computer-controlled dispensing cabinet that provides safety medication management and has attracted attention as system that can reduce errors. This system has been shown to prevent errors, such as medication mix-ups [5][6]. However, ADS requires a high cost.

Operators need to spend a lot of concentration to carefully check prescriptions and drugs on the dispensing cabinet. In order to visualize where the correct drug is stored, some methods have been proposed by using LED or projector. Han et al. proposed a method that teaches the location of drugs to operators by controlling LEDs on the cabinet with a microcontroller for notification [2]. In addition, monitoring system is developed [7]. This system visualizes shelves with projector and LEDs and measures Augmented Reality (AR) markers installed shelves for monitoring the activity to help operators find out the shelf should be operated next. However, this system needs to install LEDs or projector or AR markers on each shelf.

With recent development of computer vision and deep learning, it is possible to easily measure 3D position of human body parts from vision cameras. Martinez et al. proposed a simple Deep Neural Network (DNN) that estimated root-relative 3D joint positions based on 2D joint positions estimated from single RGB image [8]. In contrast, Moon et al. proposed a method to measure 3D joint position in real space directly from an RGB image [9]. However, the accuracy achieved by this method is not enough for judgment of drug picking activities. On the other hand, hand tracking methods have been proposed using single RGB camera. Zhang et al. proposed a real-time lightweight hand tracking model from an RGB image [10]. Their model consists of the palm detector that detects 2D bounding box of hands and the hand landmark model that detects detailed skeleton.

Azure Kinect can measure the location of human body using depth sensor [11]. The Kinect captures the range of 0.25~5.46m in 30 Frame Per Second (FPS) and the provided Body Tracking Software Development Kit (SDK) can measure the 3D joint positions. Azure Kinect Sensor SDK also enables developers to measure the location of objects measured by RGB camera installed the Kinect.

This study attempts to construct a framework for judging drug picking activities using RGB-D camera in order to detect the errors easily at low-cost. The framework judges the activity based on relative positions of operator’s hand and each shelf on dispensing cabinet. 3D hand tracking is performed from depth sensor. In addition, 2D hand landmarks

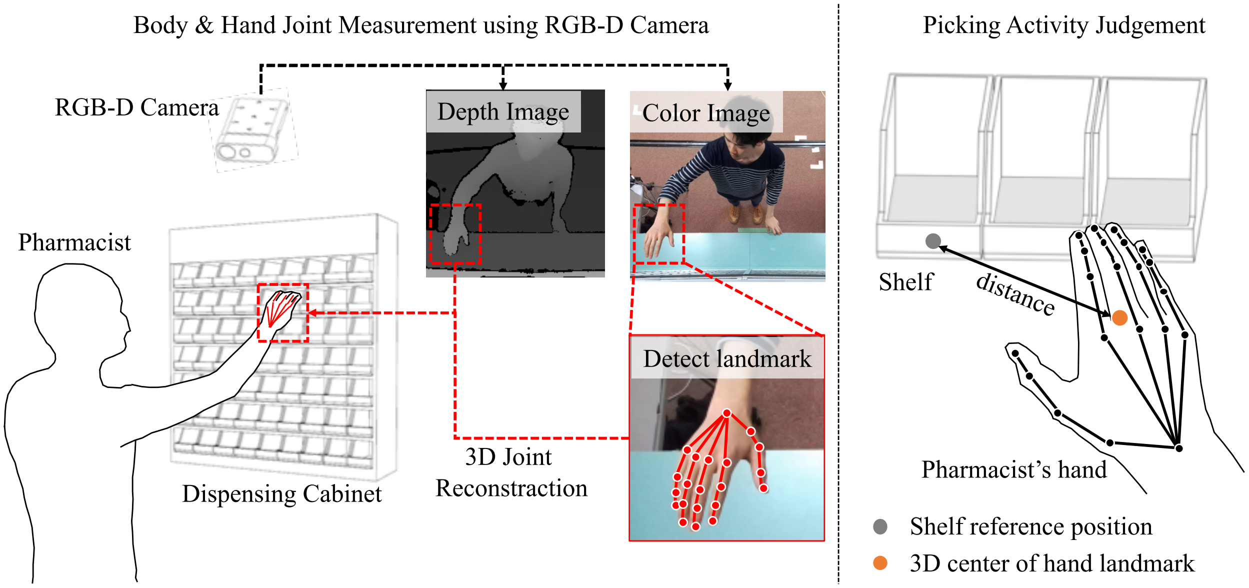
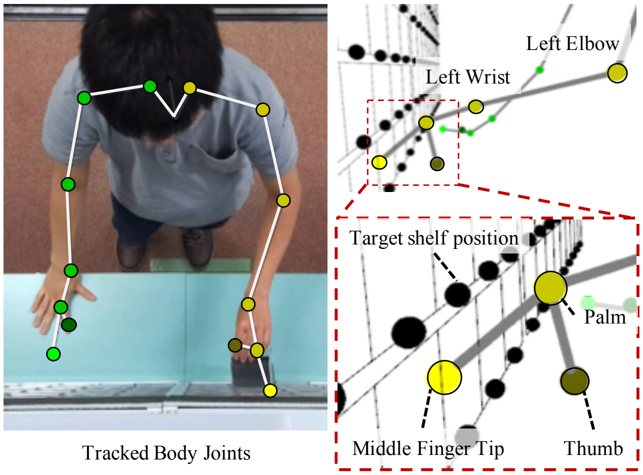
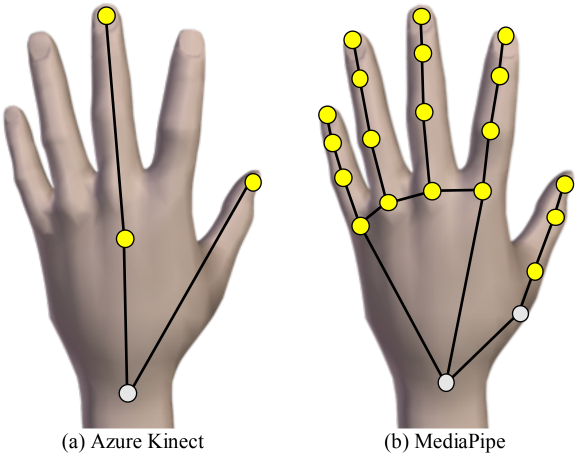


Figure 1. The illustration of drug picking activity measurement and judgement in our study.



Tracked Body Joints

Figure 2. An example of body tracking failure on Azure Kinect.


(a) Azure Kinect

(b) MediaPipe

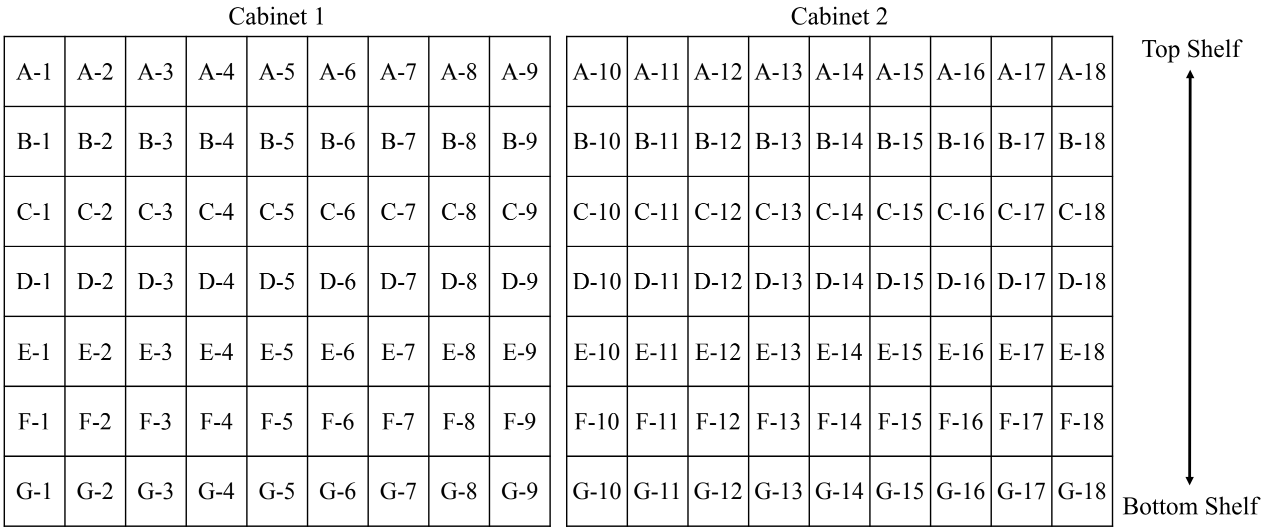
Figure 3. Hand landmarks detection in our study. Yellow landmarks are used to calculate hand center position in our study.

are tracked from RGB image using MediaPipe [10] in order to obtain stable hand measurements. 3D landmarks are inferred from the associated 2D landmarks detected by the RGB-D camera. In this study, we measure the activity of drug picking and clarify the judgement accuracy on proposed framework.

This paper is organized as follows. Section II describes the related work on human joint detection methods. Section III describes proposed methods to judge drug picking activities. Section IV describes how to assess our framework and shows result of our experiments. Section V considers about our framework's performance improvement. Finally, Section VI concludes our study.

II. Related Work

Various methods have been proposed to capture the motion of human body. The Leap Motion Controller and the Stereo IR 170 have been developed to capture hand movement [12][13]. These systems can track hands accurately, but multiple devices will be needed to measure movement during a wide range of drag picking activities. Some methods have been proposed to estimate 3D human pose from a single RGB image using DNN trained by large 3D human pose datasets [8][14][15]. Although these methods can estimate root-relative 3D location of human joints, it is difficult to directly capture their positional relationship with other object in real space. Moon et al. proposed a method to estimate the global position and posture of human based on the correlation between the size of the 2D human pose and 3D one [9].



Cabinet 1									Cabinet 2								
A-1	A-2	A-3	A-4	A-5	A-6	A-7	A-8	A-9	A-10	A-11	A-12	A-13	A-14	A-15	A-16	A-17	A-18
B-1	B-2	B-3	B-4	B-5	B-6	B-7	B-8	B-9	B-10	B-11	B-12	B-13	B-14	B-15	B-16	B-17	B-18
C-1	C-2	C-3	C-4	C-5	C-6	C-7	C-8	C-9	C-10	C-11	C-12	C-13	C-14	C-15	C-16	C-17	C-18
D-1	D-2	D-3	D-4	D-5	D-6	D-7	D-8	D-9	D-10	D-11	D-12	D-13	D-14	D-15	D-16	D-17	D-18
E-1	E-2	E-3	E-4	E-5	E-6	E-7	E-8	E-9	E-10	E-11	E-12	E-13	E-14	E-15	E-16	E-17	E-18
F-1	F-2	F-3	F-4	F-5	F-6	F-7	F-8	F-9	F-10	F-11	F-12	F-13	F-14	F-15	F-16	F-17	F-18
G-1	G-2	G-3	G-4	G-5	G-6	G-7	G-8	G-9	G-10	G-11	G-12	G-13	G-14	G-15	G-16	G-17	G-18

Figure 4. The arrangement of the dispensing cabinets with shelves and their corresponding index viewed from subject.

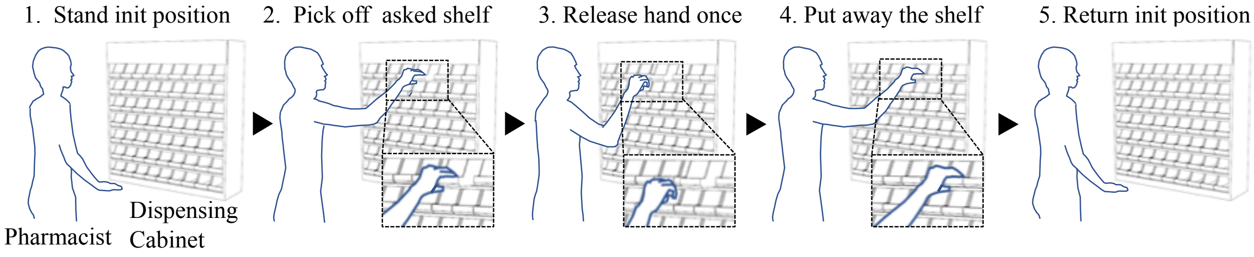


Figure 5. Drug picking activity defined in our study.

However, the measurement accuracy of this method depends on the human posture.

Azure Kinect is a device equipped with both a visible camera and depth sensor. The Kinect can directly capture human body and other objects using the depth sensor and the Body Tracking SDK. With the device-specific calibration data, the 3D position of the object can be measured from visible camera.

III. Proposed Framework

In this study, we attempt to construct a framework using RGB-D camera for judging drug picking activities to detect human errors easily at low-cost. Figure 1 shows a schematic diagram representing how to measure and judge the activity in our proposed framework. The framework uses Azure Kinect Body Tracking SDK to measure 3D position of operator’s hand from depth sensor. However, Azure Kinect may not be able to track the hand position accurately due to the pose of the human body or occlusions. Figure 2 shows an example of a body tracking failure in Azure Kinect. In order to measure the hand accurately, this study also detects hand position from RGB image by MediaPipe and estimates 3D position based on corresponding depth value from depth sensor. Finally, we determine the operated shelf based on the hand position and known shelf position. Figure 3 shows hand landmarks detection in our study.

1. 3D Body Joints Measurement using RGB-D Camera

This study places Azure Kinect above dispensing cabinet to measure the hand motion with less occlusion. The hand detection involves two steps. First, we determine the approximate position of the hand in the RGB image based on the wrist position obtained by body tracking in the Azure Kinect Sensor SDK. Next, we extract a Region of Interest (ROI) for each hand from the Kinect RGB image. These ROIs are used to detect 2D hand landmarks using the MediaPipe framework. Finally, the depth information from the Kinect's depth sensor is added to generate 3D hand landmarks.

2. Drug Picking Activity Judgement

The procedure for the decision algorithm is as follows. First, we calculate the hand position as the center position of the detected 3D hand landmarks as shown in Figure 3(b). This calculation enables a stable measurement of the hand position, especially when fingertips are hidden during the activity. Next, the distance between the hand position and each shelf position is calculated. Then, we identify the closest shelf from the hand. if both hands can be detected, we choose the shelf with the shorter distance. Finally, shelves that have been detected for

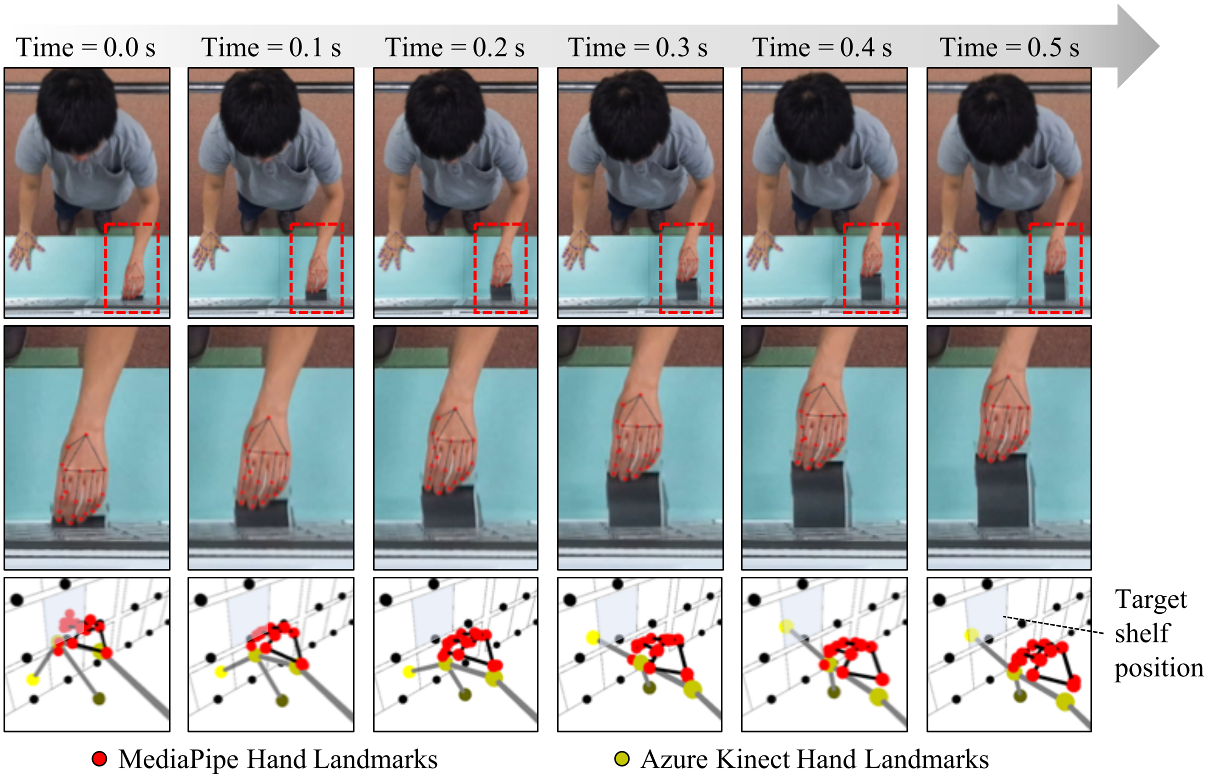


Figure 6. The result of proposed hand joint measurements in our study.

TABLE I. JUDGEMENT ACCURACY USING EACH LANDMARK.

	Landmark		
	Azure Kinect		MediaPipe
Subject	Wrist	Hand	Hand
A	43%	99%	100%
B	18%	89%	99%
C	32%	90%	99%
D	63%	93%	98%
E	30%	79%	99%
All	37%	90%	99%

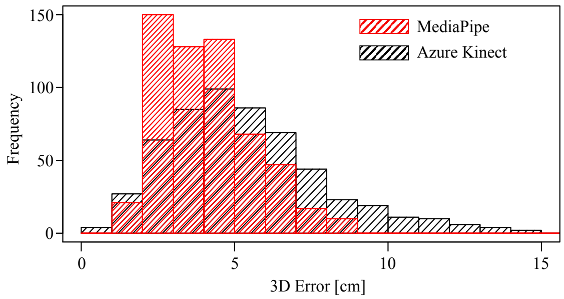


Figure 7. The Histogram of error distribution between 3D hand and target shelf when subject pick it.

more than 0.5s are judged as “operated shelf”. If more two shelves are determined, the shelf with the longer time is adopted. In this study, the drug picking activity is defined as “pulling out the target shelf”. We do not consider that whether the actual drug is obtained or not.

IV. Experiments and Results

We evaluated the proposed framework in terms of judgement accuracy and suitable hand landmarks. First, we measured the drug picking activity by subjects to verify our framework’s judgement accuracy. Second, we measured the center of each hand landmark of the Azure Kinect and the MediaPipe, and compared the distance between each hand position and the target shelf in order to clarify which hand landmark is suitable for the judgement. Figure 4 shows the arrangement of the dispensing cabinets used in this experiment and the index corresponding to each shelf. The dispensing cabinet used in this experiment can hold 63 shelves (7 rows by 9 columns). The shelf’s size is 9.4cm×10.6cm×13.3cm. We aligned two cabinets side by side and put them 85.5cm above the floor. The Azure Kinect is installed at 92cm above the cabinets. The resolution of the color camera is 1920×1080px, the field of view is 90° ×59°, the resolution of the depth sensor is 512×512px, and the field of view is 120° ×120°. We collected five healthy subjects (A~E) for the experiments.

1. Our Framework’s Activity Judgement Accuracy

The procedure of our experiments is as follows. First, we asked the subject to stand in the center of the cabinets. This position is defined as initial position. Next, we randomly

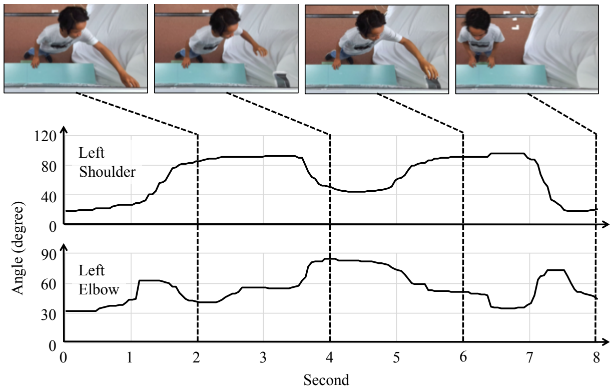


Figure 8. Left shoulder and Left elbow joint angle during subject A pick shelf, A-1.

displayed an index of shelves to choose from on a monitor set up above the shelves. Each subject is asked to pick the corresponding shelf after confirming its ID. Figure 5 shows the procedure of drug picking activities in this experiment. In order to prevent the activity from becoming uneven depending on the position of the subject at the time, we ask subject to return back the initial position after each picking action. The shelf on left side of the cabinets should be operated with the left hand, whereas that on right side of the cabinet with the right hand. When the subject performs picking, the back of the hand should face upward. Finally, we judge the activity based on our frameworks. For the experiment, the subject performs the picking activity only once per shelf. After this measurement, the experimenter confirms the activity. If subject picked a wrong shelf, this action was excluded from this evaluation.

In this experiment, we compared the picking accuracy using different human body landmark: the Azure Kinect's wrist, the center of hand landmarks of the Azure Kinect, the center of hand landmarks of the MediaPipe. Table 1 shows the resulting judgement accuracy. The hand landmarks of the MediaPipe were able to judge the activity with highest accuracy. Figure 6 shows the measurement result of the drug picking activity on shelf F-8 by subject E.

2. *Comparison of Azure Kinect hand and MediaPipe hand*

We verified whether the hand landmarks of Azure Kinect or MediaPipe are more suitable for judging the picking activity. First, we manually obtained the hand position when subject grasps target shelf. Next, we calculated 3D Euclidean distance between the hand position and the target shelf. Finally, we tested difference between the distance to the target shelf for both landmarks using Welch's t-test. A significant difference was found in the scores for the hand landmarks of MediaPipe (M=4.1cm, SD=1.5cm) and the one of Azure Kinect (M=5.3cm, SD=2.5cm); $t(902.46) = 9.95$, $p < 0.001$. There results suggest that the hand landmarks of the MediaPipe is better than Azure Kinect for the judgement. Figure 7 shows that the histogram of the 3D distance to the target shelf for the hand landmarks.

V. Discussion

In this study, we have constructed a framework to judge drug picking activities using the Azure Kinect. Compared to BCMA system and ADS, our framework does not require scanning of bar-code each time, as well as no capital investment or the high cost of installing LEDs or AR markers on dispensing cabinets. Our experiments show that our framework can accurately judge the picking activity by hand tracking at low-cost.

In order to further improve the judgement accuracy of drug picking activity, the following can be considered. First, improvement of body tracking from depth image is important to increase the detection of body joints. In this study, we have measured operator by Azure Kinect installed the top of the dispensing cabinet to capture operator's hand with less occlusion. However, the tracking will fail if the operator's body is leaning forward or closed to the cabinet. In order to solve this problem, we consider an additional training data for body tracking from the upper part of human body. Second, using the Azure Kinect's depth sensor to detect the target shelf pulled by the operator could improve the accuracy of the judgment.

Body tracking of the Azure Kinect can be used to measure the body position and joint angles of the operator. This allows us to calculate movement traveled, angular velocity of the joints during drug picking activities. Therefore, we think our framework has a potential to extract characteristic movements in the activity and the evaluation of the activity load. Figure 8 shows the joint angles of the subject A's left shoulder and elbow in shelf A-1 operation.

In order to verify effectiveness of our framework for judging drug picking activities, we enforced several constraints on the subject's movements during the experiment. In practice, however, operators can perform their drug picking activities without these constraints and use different types of shelves. In the future, we intend to evaluate the activities without such constraints, but with the Azure Kinect, we believe we can measure the reference position of each shelf of different sizes.

VI. Conclusion

In this study, we have proposed the judgement framework for drug picking activity. Our framework uses the Azure Kinect to evaluate the activity easily at low-cost. In addition to use body tracking of the Azure Kinect, we utilize hand tracking from an RGB image to track operator's hand for more stable. Our experiments show that the proposed framework enables accurately judge the activity by hand tracking taken advantage of the Azure Kinect and the MediaPipe. In the future, we will improve the accuracy of the activity judgement and measure the activity without constraints and analyze the movement of the operator.

References

[1] Pharmaceutical Safety and Environmental Health Bureau, "The state of dispensing operations", https://www.mhlw.go.jp/content/000498352.pdf [retrieved: July, 2021] (in Japanese)

[2] C. Han et al., “The assistance for drug dispensing using LED notification and IR sensor-based monitoring methods,” 2018 9th International Conference on Awareness Science and Technology (iCAST), pp. 264-267, Sept. 2018, doi:10.1109/ICAwST.2018.8517168.

[3] E. G. Poon et al., “Effect of bar-code technology on the safety of medication administration,”, The New England Journal of Medicine, Vol 362, pp. 1698-1707, May. 2010, doi: 10.1056/NEJMsa0907115

[4] J. Bonkowski et al., “Effect of barcode-assisted medication administration on emergency department medication errors,” Academic Emergency Medicine, vol. 20, pp. 801-806, Aug. 2013, doi: 10.1111/acem.12189

[5] C. Chapuis et al., “Automated drug dispensing system reduces mediaction errors in an intensive care setting,” Critical Care Medicine, Vol. 38, pp. 2275-2281, Dec. 2010, doi: 10.1097/CCM.0b013e3181f8569b

[6] J. G. Dib et al., “Effects of an automated drug dispensing system on medication adverse event occurrences and cost containment at SAMSO,” Hospital Pharmacy, Vol 41, No. 12, pp. 1180-1184, Dec. 2006, doi: 10.1310/hpj4112-1180

[7] AIOI SYSTEMS CO., LTD, “Projection Picking System,” https://www.hello-aioi.com/en/solution/digital_picking/projection/pps/ [retrieved: July, 2021]

[8] J. Martinez, R. Hossain, J. Romero, and J. J. Little, “A simple yet effective baseline for 3d human pose estimation,”, pp. 1-10, Aug. 2017, arXiv:1705.03098

[9] G. Moon, J. Y. Chang, and K. M. Lee, “Camera distance-aware top-down approach for 3d multi-person pose estimation from a single RGB-image,” pp. 1-15, Aug. 2019, arXiv:1907.11346

[10] F. Zhang et al., “MediaPipe hands: on-device real-time hand tracking,” pp. 1-5, Jun. 2020, arXiv:2006.10214

[11] Microsoft Azure, “Azure Kinect DK,” https://azure.microsoft.com/en-us/services/kinect-dk/ [retrieved: July, 2021]

[12] Ultraleap, “Leap Motion Controller, ” https://www.ultraleap.com/product/leap-motion-controller/ [retrieved: July, 2021]

[13] Ultraleap, “Ultraleap Stereo IR 170,” https://www.ultraleap.com/product/stereo-ir-170/ [retrieved: July, 2021]

[14] H. Fang, Y. Xu, W. Wang, X. Liu, and S. Zhu, “Learning Pose Grammar to Encode Human Body Configuration for 3D Pose Estimation,” The Thirty-Second AAAI Conference on Artificial Intelligence, pp. 6821-6828, Jan. 2018, arXiv:1710.06513

[15] X. Chen, K. Lin, W. Liu, C. Qian, and L. Lin, “Weakly-Supervised Discovery of Geometry-Aware Representation for 3D Human Pose Estimation,” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10895-10904, Jun. 2019