3D Human Pose Estimation using a Stereo Camera towards Monitoring of Drug Picking Tasks

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

***Abstract*—Medication dispensing errors are a critical issue that threatens the quality of health care services and patients' safety. Checks by pharmacists are effective in preventing this potential problem but increase the workload. In Japan, dispensing assistants are allowed to assist pharmacists in picking drugs to reduce their workload. However, becoming a dispensing assistant requires experiences, as drugs can easily be mistaken because of their similar names and forms. This study attempts to construct a monitoring framework using stereo camera-based Three-Dimensional (3D) human pose estimation for detecting errors and evaluating the physical workload during the drug picking task. To accurately estimate the 3D human pose, we improve the estimation of body joints and perform calibration using multiple reference points. Our results show that the proposed framework can estimate the 3D human pose with acceptable accuracy to detect errors in the drug picking task. Future work will examine the applicability of the proposed framework to the assessment of physical workload of dispensing assistants.**

Keywords-3D human pose estimation; Stereo camera; Medication dispensing error; Pharmacy.

I. INTRODUCTION

As the global elderly population grows, the proportion of patients with multimorbidity will increase [1]. Patients with multimorbidity are prescribed more medications and are therefore at higher risk for medication dispensing errors. Such errors are serious threats to the quality of health care services and a patients' safety [2].

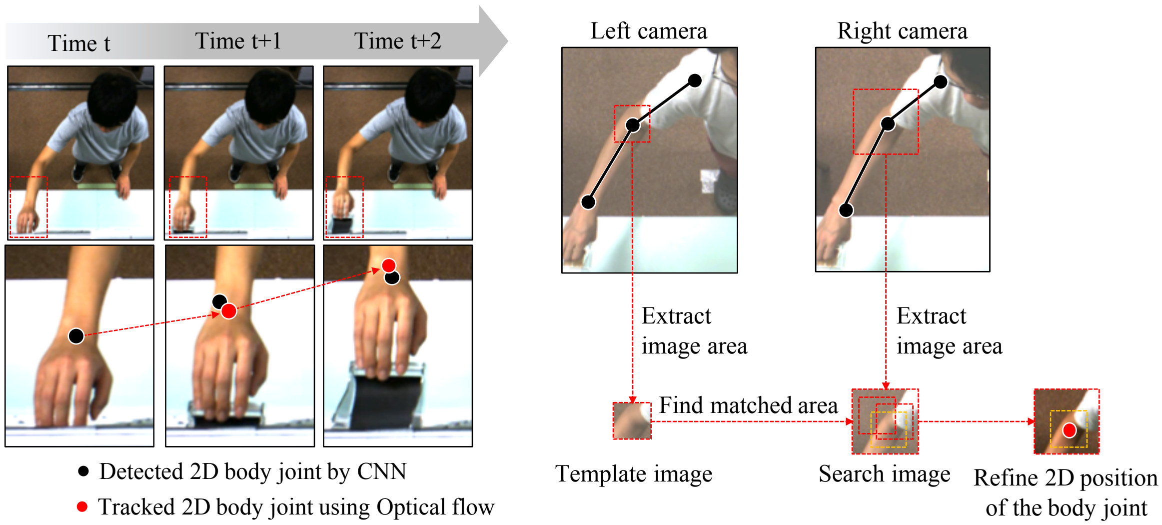
The presence of pharmacists is essential to prevent medication dispensing errors. In addition to dispensing and supplying prescription drugs, pharmacists are expected to apply their specialized knowledge and skills to a variety of tasks, such as detecting medical side effects and providing medication guidance [3][4]. However, workload of pharmacists is increasing as their scope of work expands [5]. Previous study has shown that the high workload may lead to overlook the risk of health problems in patients [6]. High physical workload including prolonged standing, sitting and repetitive tasks may also cause musculoskeletal disorders (MSDs) of the body, such as neck and back [7].

In Japan, Pharmaceutical Safety and Environmental Health Bureau has allowed dispensing assistants to pick drugs starting April 2019 to reduce workload of pharmacists and maintain the quality of healthcare services [8]. This expects that pharmacists focus on more specialized activities. However, the task is prone to cause dispensing errors because the names and the forms of drugs are similar. In addition, pharmacists must still check the dispensed drug eventually. Various methods have been proposed to prevent the dispensing error [9][10]. However, these methods require an operator to operate the cumbersome process, such as scanning a bar-code or modifying each shelf. In addition, it is difficult for these methods to evaluate the physical workload from operator's movements during the picking task.

Three-Dimensional (3D) human pose estimation using vision cameras can measure 3D movements of human body without contact. Especially, methods using multiple vision cameras can measure the position of the body joint relative to surrounding objects. However, if there is a discrepancy between the Two-Dimensional (2D) correspondence position of the body joint estimated from each camera image, it may cause large errors in its resulting 3D position [11].

This study attempts to construct a monitoring framework based on stereo camera-based 3D human pose estimation for detecting errors in drug picking tasks and evaluating the physical workload of dispensing assistants during the tasks. We propose two methods to estimate 3D human pose more accurately. First method is a 2D body joint position correction method, which estimate plausible 2D position of a certain body joint in each camera image. Second method is a 3D calibration method using multiple reference points to refine the 3D measurement accuracy by the stereo camera. In addition, this study introduces a method to determine the correct picking task based on the estimated 3D position of the hand. Finally, we will verify accuracies of the 3D human pose estimation and of the picking task. We will also discuss whether this framework is applicable to the evaluation of the physical workload of the operator.

This paper is organized as follows. Section II describes the related work on prevention of the error during drug picking tasks and 3D human pose estimation from vision cameras. Section III describes the proposed framework to estimate accurate 3D human pose and determine the picking task. In Section IV, we clarify the 3D human pose estimation accuracy and the accuracy of the picking task. Section V consider about performance improvements and evaluation of physical workload in this study. Finally, Section VI concludes our study and describes future works.



(a) Optical flow-based correction (b)Template matching-based correction

Figure 1. 2D body joint correction on a stereo image

II. Related Work

There are various methods proposed to prevent errors in drug picking tasks. Barcode-Assisted Medication Administration (BCMA) system manages drugs, prescriptions, and patient information by reading corresponding barcodes [9][12]. In addition, medical dispensing system using RFID is proposed [13]. However, these systems are complicated because they require to read barcodes or tags during the picking task. Few methods using Light Emitting Diode (LED) or Augmented Reality (AR) markers are proposed [10][14]. However, these methods are also cumbersome because we need to modify dispensing cabinets to put LEDs or AR markers.

The technology of 3D human pose estimation has made significant progress. Azure Kinect can estimate the 3D position of body joints using the Body Tracking Software Development Kit (SDK) from depth data measured by a depth sensor [15]. Ono and Prima indicated the possibility to monitor the drug picking task by utilizing 3D hand detection combining MediaPipe [16] and Azure Kinect [17]. However, this method can only measure the picking task within the range that can be measured by the Azure Kinect.

There are also methods using visible light cameras to estimate the 3D position of body joints. These methods can be roughly classified into methods using single camera and methods using multiple cameras. Methods using single camera utilize deep neural network to estimate 3D human pose from a camera view [18][19]. However, it is difficult for these methods to measure the accurate 3D position of the body joint and other objects in real space due to the direction of the body and self-occlusion and so on. Methods using multiple cameras have advantageous that these can calculate 3D position of objects using triangulation. Nakano et al. used the Convolutional Neural Network (CNN) [20] to estimate the 2D position of body joints and estimated the 3D position of the body joint by triangulation. However, this method needs accurate the 2D position of body joints for accurate estimation of the 3D position of body joints with multiple cameras [11]. Sayo et al. proposed a refinement network that estimates the difference between the 2D position of body joints estimated using a CNN and the 2D projection of the actual 3D position of body joints [20]. This method can estimate the accurate the 2D position of body joints by subtracting the difference the network estimated. However, this method doesn't consider the temporal consistency of the 2D position of body joints or the discrepancy of the 2D correspondence position of body joints estimated on multiple camera images.

III. Proposed Framework

We installed a stereo camera on the top of the dispensing cabinet to capture drug picking tasks. Our framework detects the 2D position of body joints on each camera image using CNN [21]. The resulting 2D body joints are corrected to identify plausible 2D position for the same body joint on each camera image. The 3D position of body joints is estimated based on triangulation method and optimized by a 3D calibration method using multiple reference points. Finally, our framework determines the picking task of the operator using the estimated 3D hand position.

A. 2D Body Joints Position Tracking Using Optical Flow

In this study, we utilize optical flow to track the 2D position of body joints consistently in time. Figure 1(a) shows correction of 2D position of body joints using optical flow. Because the tracking using only optical flow may fail due to occlusion, we set multiple tolerance levels on the distance between the 2D position of body joints tracked by the optical flow and the 2D position of body joints detected by the CNN. We then define new 2D position of body joints according to the tolerance levels. Our body joint tracking is as follow. The 2D position of the body joint $p_t(x_j, y_j)$ detected by the CNN at time t is defined as an initial position of the body joint. Next, we track the 2D position of the body joint $p'_{t+1}(x'_j, y'_j)$ on the image at time $t+1$ using optical flow based on the Lucas-Kanade method [22]. We calculate the distance between the

 ISBN: 978-1-61208-982-9

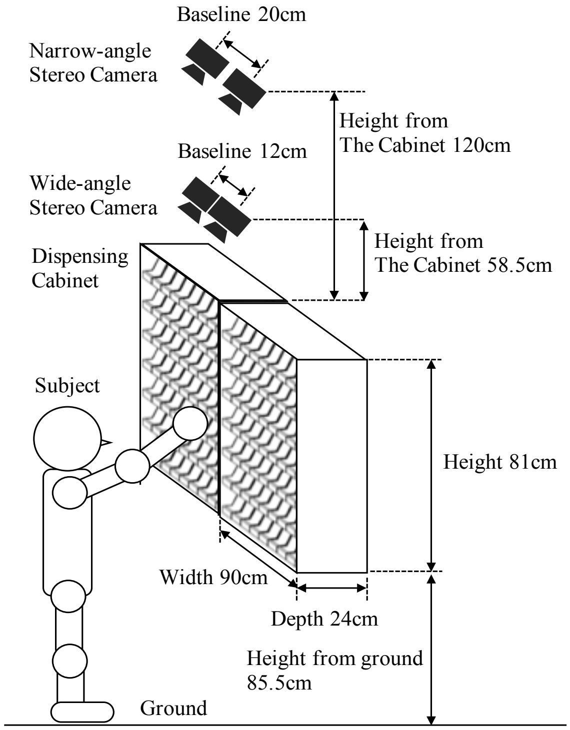


Figure 2. Setup of stereo cameras and dispensing cabinets

2D position of the body joint $p_{t+1}(x_j, y_j)$ detected by CNN and the 2D position of the body joint $p'_{t+1}(x'_j, y'_j)$ tracked using optical flow. Finally, new 2D position of the body joint is determined according to the distance. In this study, $p'_{t+1}(x'_j, y'_j)$ is adopted if the distance is less than 5px. If the distance is greater than 5 pixels and less 10 pixels, the midpoint of both points is adopted. If the distance is greater than 10 pixels, $p_{t+1}(x_j, y_j)$ is adopted.

B. 2D Body Joint Correction on a Stereo Image Using Template Matching

To identify plausible 2D position for same body joint across stereo images, this study corrects the 2D position of the body joint between two images by finding plausible the position using template matching. Figure 1(b) shows illustration of method for correction of the 2D position of the body joint. At first, we generate a template image and a search image. These images are within a certain range from the body joint detected in each image. Next, we find the plausible 2D position of the body joint by searching for the area of highest similarity to the template images. Finally, the center of the area is defined as the corrected 2D position of the body joint. In this study, the size of the template image is 16×16 pixels, and the search range is 33×33 pixels.

C. 3D Calibration Using Multiple 3D Reference Points

The 3D position of body joints cannot be accurately estimated due to camera lens distortion. To solve this problem, we propose a 3D calibration method using multiple reference points. The reference points are placed entirely in the field of view of a stereo camera. Next, the 3D position of them is estimated based on triangulation. The actual measured position of the reference point $P_i = (X_i, Y_i, Z_i)$ and the estimated position $P'_i = (X'_i, Y'_i, Z'_i)$ is equal to

$$P_i = A \cdot P'_i + \zeta. \tag{1}$$

where A is $m \times n$ refinement matrix, ζ is residuals. The matrix A is calculated by least squares method.

$$\text{argmin} \sum_{i=1}^{N} \|A \cdot P'_i - P_i\|. \tag{2}$$

The final 3D position of the body joint $J'_j = (X'_j, Y'_j, Z'_j)$ is defined by

$$J'_j = A \cdot J_j. \tag{3}$$

where J_j is 3D position of the jth body joint before the application of the 3D calibration method.

D. Picking Task Evaluation Based on 3D Hand Position

The method for evaluating drug picking tasks based on the 3D position of the operator's hand is as follows. First, the 3D position of the hand is calculated as the center position of 3D hand landmarks estimated by our framework. Next, the distance between the hand position and each shelf position is calculated to determine the closest shelf. Finally, when the shelf and hand positions overlap for more than 0.5 seconds, the operator is considered to have operated the shelf.

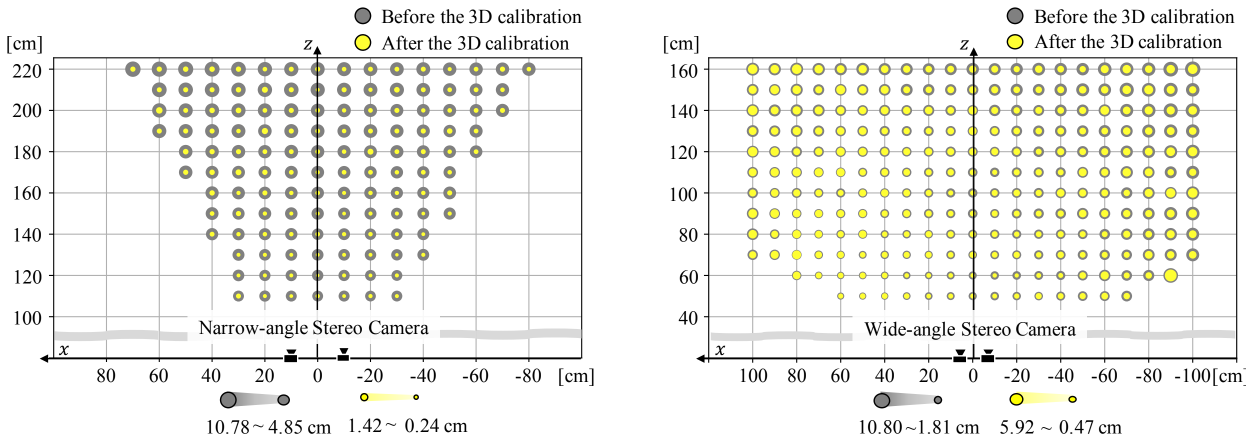
IV. EXPERIMENTS AND RESULTS

We verify the performance of the proposed framework towards monitoring of the drug picking task. First, we calculate the measurement error of the 3D position of multiple reference points placed in the field of view of the stereo camera to clarify the 3D measurement accuracy. Next, we capture the picking task by subjects installed few markers to verify the estimation accuracy of the 3D human pose. Finally, we also verify whether the proposed framework can determine the correct picking task based on the 3D hand position.

A. Experimental Environment

Figure 2 shows the setup of stereo cameras and dispensing cabinets. The experimental room is 300cm high from the floor to the ceiling. In this experiment, we use the dispensing cabinet that can store 63 shelves. The size of a shelf is 9.4cm×10.6cm×13.3cm. Two dispensing cabinets are placed side by side. Two types of stereo cameras are used in this experiment: the MD-SUA133GC-T and the CaliCam, manufactured by Shenzhen MindVision Technology Co. Ltd. and Astar.ai Inc., respectively. Each camera in the MD-SUA133GC-T has a 48.5° x 36.9° field of view and captures images at 60 Frames Per Second (FPS). The baseline length between cameras is 20cm. The CaliCam stereo camera has a field of view of 120° x 100° at 30 FPS and the baseline length is 12 cm. Hereinafter, MD-SUA133GC-T is referred to as a narrow-angle stereo camera and CaliCam as a wide-angle stereo camera.

ISBN: 978-1-61208-982-9



(a) Narrow-angle Stereo Camera (b) Wide-angle Stereo Camera

Figure 3. Measurement error of 3D reference points

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	Top
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	Bottom

74 Shelves can be measured by narrow-angle stereo camera

Figure 4. Location of shelves used in this study

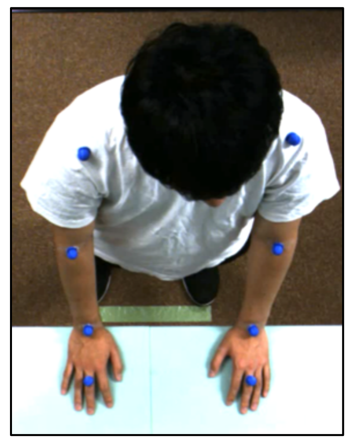


Figure 5. Location of markers

B. 3D Point Measurement Accuracy by a Stereo Camera

First, reference points were placed in the field of view of each stereo camera. These points for correction were regularly placed at 10 cm intervals. Next, these 2D positions were obtained from each camera image and the corresponding 3D positions were measured. Finally, the difference between the estimated and actual 3D positions was calculated in Root Mean Square Error (RMSE).

Figure 3 shows the measurement error of the reference points. By using the narrow-angle stereo camera, the RMSE of all reference points was 7.8 cm before and 0.8 cm after the application of the 3D calibration. Whereas, the RMSE of all reference points was 5.0 cm before and 1.8 cm after the application of the 3D calibration when a wide-angle stereo camera was used. As a result, we found that the 3D calibration method can improve the 3D measurement accuracy regardless of the angle of view of the stereo camera.

C. The Accuracy of 3D Human Pose Estimation

To verify the accuracy of the 3D human pose estimation, markers were attached to the body of the operator doing the picking task, and the difference between the position of markers and the position of body joints estimated by the proposed framework was calculated. Six shelves near the center of the two cabinets (A-6, D-6, G-6, A-13, D-13, G-13) and four shelves on either side of the cabinet (A-1, A-18, G-1, G-18) were used for the picking task. Figure 4 shows the location of shelves used in this study. The wide-angle stereo camera can measure all shelves, whereas the narrow-angle stereo camera can measure only 74 shelves. Four subjects participated in this experiment.

The procedure for the evaluation of the 3D pose estimated by our framework is as follows. Figure 5 shows the location of markers. We put markers on subject's shoulders, elbows, wrists, and hands. The marker is a blue sphere with a diameter of 1.8cm. First, the subject stands in the center of the two cabinets and performs the picking task. Next, the 2D position of each marker is extracted by using HSV color extraction and blob detection. Then, the 3D position of each marker is calculated by triangulation and refine the position by the 3D calibration method. Finally, the error between the 3D position of body joints estimated by our framework and the position of the corresponding markers is calculated as RMSE and Standard Deviation (SD). We calculate the hand position as the center position of the detected 3D hand landmark by our framework.

Tables I and II show the accuracy of the 3D human pose estimation during the picking task for shelves that can be measured both stereo cameras. Table III shows the accuracy of the 3D human pose estimation during the picking task for

TABLE I. ACCURACY OF 3D HUMAN POSE ESTIMATION USING THE NARROW-ANGLE STEREO CAMERA DURING DRUG PICKING TASK [CM]

Body Joint		Methods			
		None RMSE (SD)	Optical Flow RMSE (SD)	Template Matching RMSE (SD)	Optical Flow and Template Matching RMSE (SD)
Right	Shoulder	4.8 (2.1)	4.8 (2.1)	2.5 (1.0)	2.4 (0.9)
	Elbow	5.6 (2.7)	5.5 (2.5)	2.8 (1.6)	2.8 (1.6)
	Wrist	4.8 (2.1)	4.5 (2.0)	1.6 (0.7)	1.6 (0.7)
	Hand	3.1 (0.9)	3.0 (0.9)	2.1 (0.8)	2.2 (0.8)
Left	Shoulder	4.3 (2.1)	4.2 (2.1)	2.0 (0.9)	1.9 (0.8)
	Elbow	4.9 (2.4)	5.1 (2.4)	1.9 (0.7)	1.8 (0.7)
	Wrist	4.4 (2.1)	4.1 (1.9)	2.0 (0.8)	2.0 (0.8)
	Hand	3.0 (1.1)	3.0 (1.1)	2.1 (0.8)	2.1 (0.8)
Mean of RMSE		4.36	4.28	2.13	2.10

TABLE II. ACCURACY OF 3D HUMAN POSE ESTIMATION USING THE WIDE-ANGLE STEREO CAMERA DURING DRUG PICKING TASK [CM]

Body Joint		Methods			
		None RMSE (SD)	Optical Flow RMSE (SD)	Template Matching RMSE (SD)	Optical Flow and Template Matching RMSE (SD)
Right	Shoulder	8.3 (3.6)	7.6 (3.4)	3.7 (1.3)	3.5 (1.2)
	Elbow	6.5 (3.1)	5.9 (2.8)	3.6 (1.7)	3.6 (1.7)
	Wrist	8.1 (4.5)	8.4 (4.9)	2.8 (1.3)	2.9 (1.3)
	Hand	3.8 (1.5)	3.8 (1.5)	2.6 (1.1)	2.6 (1.1)
Left	Shoulder	7.4 (3.7)	6.3 (3.3)	2.5 (1.1)	2.4 (1.0)
	Elbow	6.5 (3.6)	6.2 (3.5)	2.8 (1.1)	2.8 (1.1)
	Wrist	6.6 (3.4)	7.0 (3.9)	2.8 (1.1)	2.9 (1.2)
	Hand	4.2 (1.8)	4.2 (1.8)	2.6 (0.9)	2.6 (0.9)
Mean of RMSE		6.43	6.18	2.93	2.91

TABLE III. ACCURACY OF 3D HUMAN POSE ESTIMATION USING THE WIDE-ANGLE STEREO CAMERA DURING WIDE-AREA DRUG PICKING TASK [CM]

Body Joint		Methods			
		None RMSE (SD)	Optical Flow RMSE (SD)	Template Matching RMSE (SD)	Optical Flow and Template Matching RMSE (SD)
Right	Shoulder	8.6 (4.1)	7.5 (3.6)	3.9 (1.7)	3.8 (1.7)
	Elbow	7.6 (4.0)	6.9 (3.9)	3.6 (1.5)	3.5 (1.5)
	Wrist	8.1 (4.2)	7.8 (4.2)	3.3 (1.4)	2.9 (1.2)
	Hand	4.8 (2.2)	5.0 (2.4)	3.2 (1.5)	3.0 (1.4)
Left	Shoulder	6.9 (3.5)	5.7 (3.2)	3.2 (1.5)	3.0 (1.5)
	Elbow	7.0 (3.7)	6.6 (3.5)	3.4 (1.2)	3.4 (1.1)
	Wrist	7.5 (3.9)	7.8 (4.2)	3.1 (1.3)	3.0 (1.2)
	Hand	5.3 (2.4)	5.6 (2.6)	3.5 (1.5)	3.3 (1.5)
Mean of RMSE		6.98	6.61	3.40	3.24

shelves that can be measured only the wide-angle stereo camera. These results show the accuracy of 3D human pose estimation from a wide-angle stereo camera is equivalent to that from a narrow-angle stereo camera. Table IV shows the result of Tukey-Kramer multiple comparison test on the mean of RMSE of 3D human pose estimation improved using optical flow and template matching. A significant difference was found in mean score of RMSE of the 3D human pose estimated using template matching and not. However, there was not a statistically significant difference in mean score of RMSE of the 3D human pose estimated using optical flow.

D. *The Determination Accuracy of Picking Task*

This experiment confirms whether the proposed framework can determine the correct picking task based on the estimated 3D position of the hand. Subjects are randomly instructed which shelf to operate. Pulling a shelf and returning it to its original position is defined as a picking task. The left-hand shelf is operated by the left hand whereas the right-hand shelf is operated by the right hand. In addition, when the subject operates the shelf, the back of the hand should face upward. In this experiment, we first captured the picking task for 74 shelves that could be measured from both stereo cameras to compare the determination accuracy of the picking task. Then, we also captured the picking task for all shelves using only the wide-angle stereo camera. Four subjects participated in this experiment.

Table V shows the determination accuracy of the picking task in this experiment. The results show that the proposed framework enabled an accurate assessment of the picking task. Measurements with the wide-angle stereo camera were more accurate. The narrow-angle stereo camera is mounted higher than the wide-angle stereo camera. This results in measuring the subject from a more extreme angle, and if the subject is leaning forward, the camera may not be able to detect the subject in the image.

 ISBN: 978-1-61208-982-9

TABLE IV. TUKEY-KRAMER MULTIPLE COMPARISONS TEST FOR THE ACCURACY OF 3D HUMAN POSE ESTIMATION

Group 1	Group 2	Experiment Setting		
		Using the NA stereo camera	Using the WA stereo camera	
		Drug Picking Task p-value	Drug Picking Task p-value	Wide-area Drug Picking Task p-value
None	Optical Flow	.994	.974	.840
None	Template Matching	< .001	< .001	< .001
None	Optical Flow and Template Matching	< .001	< .001	< .001
Optical Flow	Template Matching	< .001	< .001	< .001
Optical Flow	Optical Flow and Template Matching	< .001	< .001	< .001
Template Matching	Optical Flow and Template Matching	.999	.999	.982

NA:"Narrow-angle", WA:"Wide-angle"

TABLE V. DETERMINATION ACCURACY OF DRUG PICKING TASK

Subject	Stereo Camera		
	Narrow-angle	Wide-angle	
	74 shelves	74 shelves	All shelves
A	91.9 %	95.9 %	92.9 %
B	93.2 %	100.0 %	93.7 %
C	100.0 %	100.0 %	95.2 %
D	95.9 %	100.0 %	96.0 %
All	95.3 %	99.0 %	94.4 %

V. DISCUSSION

In this study, we attempted to determine which shelf was operated by the operator using the proposed framework. The result shows the proposed framework can accurately determine which shelf was operated during the picking task. Although the proposed framework can measure a wider area than the Azure Kinect, the determination accuracy of picking task is lower than that of Ono et al. [17]. For more accurate determination, it would be possible to use the 3D position of the pulled-out shelf in addition to the 3D position of the operator's hand. This would allow detection of errors during picking tasks where the 3D human pose could not be estimated by the stereo camera. In future, we plan to improve the determination method of picking tasks and clarify the accuracy on experiments with more subjects.

The proposed framework allows to measure not only the picking task but also other tasks, such as storing drugs into the cabinet, packings of drugs or syrups. Since our framework can calculate the angle of body joints or the body movement based on the estimated 3D human pose, it might be possible to visualize and evaluate the physical workload from this data. This attempt will be useful to improve the working environment or the workflow in pharmacies.

VI. CONCLUSION

This study attempted to construct a monitoring framework based on 3D human pose estimation using a stereo camera for detecting errors in drug picking tasks and evaluating the physical workload during the picking task. Our framework utilized optical flow and template matching to accurately estimate the 2D body joints. A 3D calibration method using multiple 3D reference points was also introduced for more accurate estimation of 3D positions. Our results show that accuracy of the 3D human pose estimated by the framework was acceptable to determine whether the operator performed the picking task properly or not. Furthermore, the framework can be applied to the evaluation of the physical workload on the operator. Future work will include extending the framework to monitor the actual drug picking tasks.

REFERENCES

[1] A. H. Lavan, P. F. Gallagher, and D. O'Mahony, "Methods to reduce prescribing errors in elderly patients with multimorbidity," Clinical Interventions in Aging, vol. 11, pp. 857-866, Jun. 2016, doi:10.2147/CIA.S80280

[2] G. P. Velo and P. Minuz, "Medication errors: prescribing faults and prescription errors," British Journal of Clinical Pharmacology, vol. 67, No. 6, pp. 624-628, Jun. 2009, doi: 10.1111/j.1365-2125.2009.03425.x

[3] J. L. Schnipper et al, "Role of pharmacist counseling in preventing adverse drug events after hospitalization," Archives of Internal Medicine, vol. 166, No. 5, pp. 565-71, Mar. 2006, doi: 10.1001/archinte.166.5.565

[4] M. D. Murray, M. E. Ritchey, J. Wu, and W. Tu, "Effect of a pharmacist on adverse drug events and medication errors in outpatients with cardiovascular disease," Archives of Internal Medicine, vol 169, No. 8, pp. 757-63, Apr. 2009, doi:10.1001/archinternmed.2009.59

[5] S. Shao et al, "Workload of pharmacists and the performance of pharmacy services," PLoS ONE, vol. 15, No. 4, pp. 1-12, Apr. 2020, doi:10.1371/journal.pone.0231482

[6] D. C. Malone et al, "Pharmacist workload and pharmacy characteristics associated with the dispensing of potentially clinically important drug-drug interactions," Medical Care, vol. 45, No. 5, pp. 456-462, May. 2007, doi:10.1097/01.mlr.0000257839.83765.07

[7] A. D. Nasution and E. L. Mahyuni, "The impact of work method on musculoskeletal disorders complaints in pharmacy unit," Disease Prevention and Public Health Journal, vol. 14, No. 2, pp. 81-89, Sept. 2020, doi:10.12928/dpphj.v14i2.2478

[8] Pharmaceutical Safety and Environmental Health Bureau, "The state of dispensing operations", https://www.mhlw.go.jp/content/000498352.pdf [retrieved: June, 2022] (in Japanese)

[9] 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

[10] AIOI SYSTEMS CO., LTD, "Projection Picking System," https://www.hello-aioi.com/en/solution/digital_picking/projection/pps/ [retrieved: June, 2022]

[11] N. Nakano et al., "Evaluation of 3D markerless motion capture accuracy using OpenPose with multiple video cameras," Front. Sports Act. Living, vol 2, pp. 1-9, May. 2020, doi:10.3389/fspor.2020.00050

[12] 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

[13] S. C. Shieh, C. C. Lin, T. F. Yang, and G. H. Tu, "Using RFID technology on clinic's pharmacy operation management and development of intelligent medicine dispensing cabinet," 2008 IEEE International Conference on Industrial Engineering and Engineering Management, pp. 2006-2009, Dec. 2008, doi: 10.1109/IEEM.2008.4738223

[14] 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

[15] Microsoft Azure, "Azure Kinect DK," https://azure.microsoft.com/en-us/services/kinect-dk/ [retrieved: June, 2022]

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

[17] Y. Ono and O. D. A. Prima, "Assessment of drug picking activity using RGB-D camera," The Fourteenth International Conference on Advances in Computer-Human Interactions, ACHI2021, pp. 6-11, 2021

[18] J. Martinez, R. Hossain, J. Romero, and J. J. Little, "A simple yet effective baseline for 3d human pose estimation," Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 2640-2649, Oct. 2017

[19] 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," Proceeding of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 10133-10142, Oct. 2019

[20] A. Sayo, D. Thomas, H. Kawasaki, Y. Nakashima, and K. Ikeuchi, "PoseRN: A 2D Pose Refinement Network For Bias-Free Multi-View 3D Human Pose Estimation," 2021 IEEE International Conference on Image Processing (ICIP), pp. 3233-3237, Sept. 2021, doi:10.1109/ICIP42928.2021.9506022

[21] Z. Cao, G. Hidalgo, T. Simon, S. Wei, and Y. Shikh, "OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields," arXiv preprint, pp. 1-14, 2018, arXiv:1812.08008v2

[22] B. D. Lucas and T. Kanade, "An interative image registration technique with an application to stereo vision," Proc. the 7th international joint conference on Artificial intelligence, vol. 2, pp. 674-679, Aug. 1981, doi:10.5555/1623264.1623280