岩手県立大学
ソフトウェア情報学部
Faculty of Software and Information Science

Iwate Prefectural University

ACHI 2022, The 15th International Conference on Advances in Computer-Human Interactions
PAVIP: Practical Applications of Visual Information Processing

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

Yuta Ono and Oky Dicky Ardiansyah Prima
ACHI2022 | June 2022
g236s001@s.iwate-pu.ac.jp

IARIA

■ About Me

- **Name**: Yuta Ono
- **Course**: Ph.D. candidate student
- **Affiliation:** Graduate School of Software and Information Science, Iwate Prefectural University
- **Research of interest**
  - 3D human pose estimation and its application
  - Human activity recognition
  - Human behavior analysis

Agenda

- Background
- Research Aim
- Our Framework
- Experiments
- Picking Task Determination:
    - The narrow-angle stereo camera
    - The wide-angle stereo camera
- Conclusion

■ Background

- Medication dispensing errors are a critical issue that threatens the safety of patients’ health.
- Pharmacists are expected to apply their specialized knowledge and skills on a variety of tasks.

<u>Pharmacists’ scope of work and its effectiveness</u>

- Dispensing drugs
- Providing medication guidance to patients
- Detecting medical side effects
- Contributing their knowledge to patients’ therapy

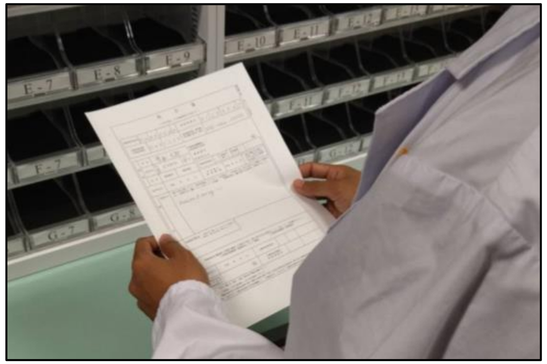
Contribute to fast and effective treatment of patients and maintenance of their health.

→ Increasing the workload of the pharmacist may result in missing errors that could be detected and their physical disability.

■ Background

- Dispensing assistants are allowed to pick drugs to reduce the workload of pharmacists in Japan.

<u>What is the drug picking task?</u>



1. Check the prescription

2. Pick the drug

3. Check the drug

4. Explain to the patient

→ Drug picking task is prone to cause dispensing errors because the names and the forms of drugs are similar.

Iwate Prefectural University

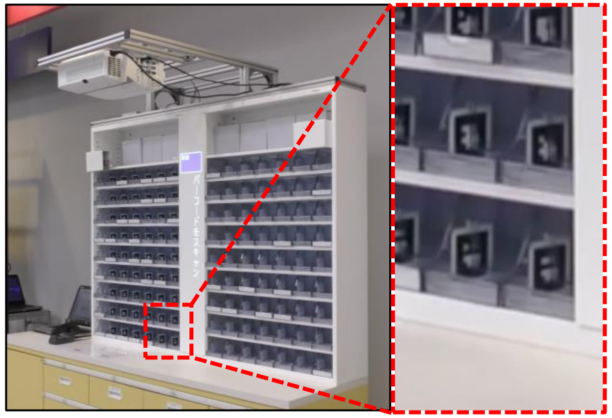
■ Background

<u>Current methods to prevent the dispensing error</u>

- **Method using barcodes or RFIDs**
  - Verify the consistency of picking tasks by scanning them
  - Pros : Easy to use
  - Cons: Require to scan them every time
- **Automate Dispensing Cabinets (ADCs)**
  - Pick correct drugs automatically
  - Pros : Precise picking
  - Cons: Occupy large space
- **Dispensing Cabinet Modification (DCM)**
  - Determine the operated shelf by detecting the AR marker
  - Pros : Detect the operation automatically
  - Cons: Cumbersome to modify each shelf

Barcode scanner [1]

ADCs [2]



DCM using AR marker [3]

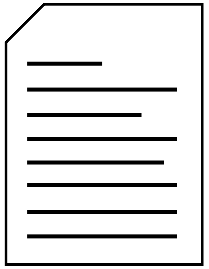
Existing methods don't consider the physical burden of the operator during the task.

Iwate Prefectural University

■ Research aim

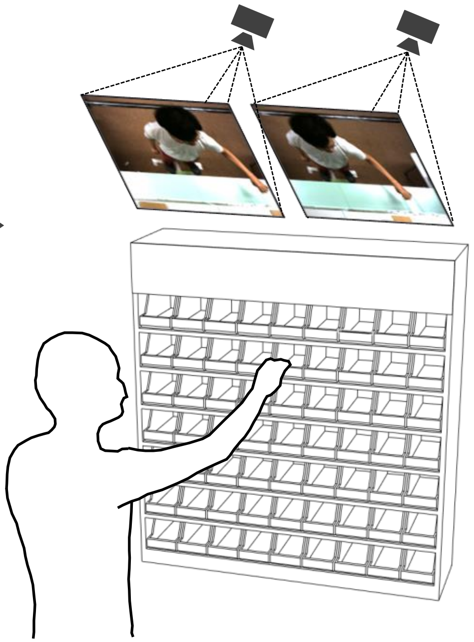
This study attempts to construct a monitoring framework based on stereo camera-based 3D human pose estimation for detecting the dispensing error and evaluating the physical workload of the operator.

1. Get the prescription

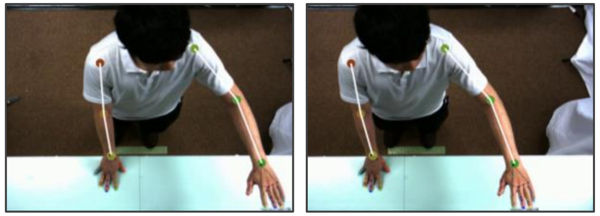


2. Picking task measurement

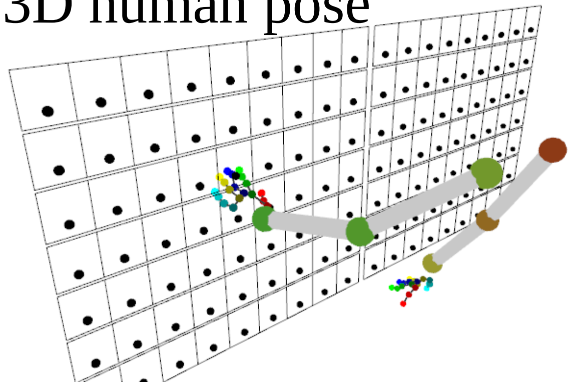
Stereo camera



Stereo image

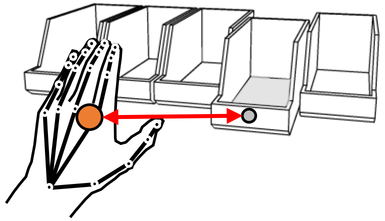


3D human pose

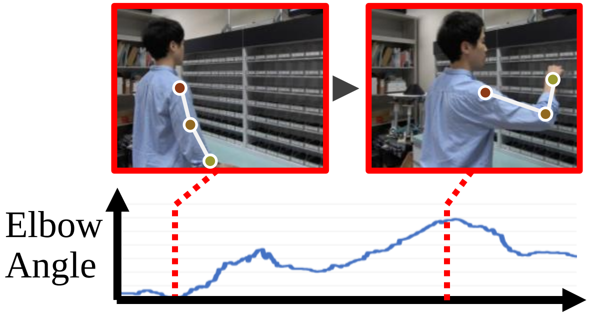


3. Task evaluation

Operated shelf detection



Physical burden analysis



■ Our framework: **Stereo camera-based 3D human pose estimation**

2D body joint detection
from a stereo image using CNN[4]

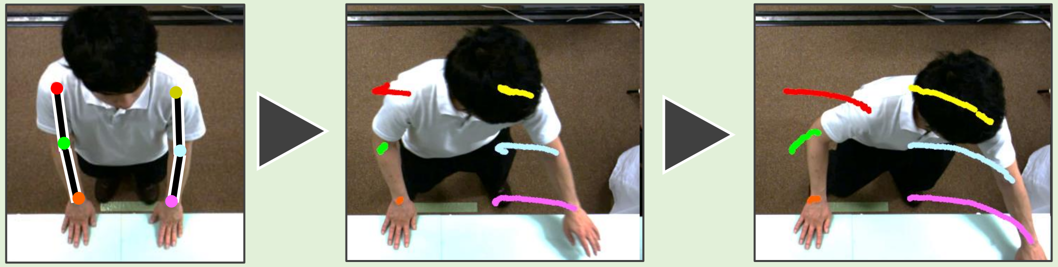
2D body joint position tracking
by optical flow

Correction of 2D body joint position
using template matching

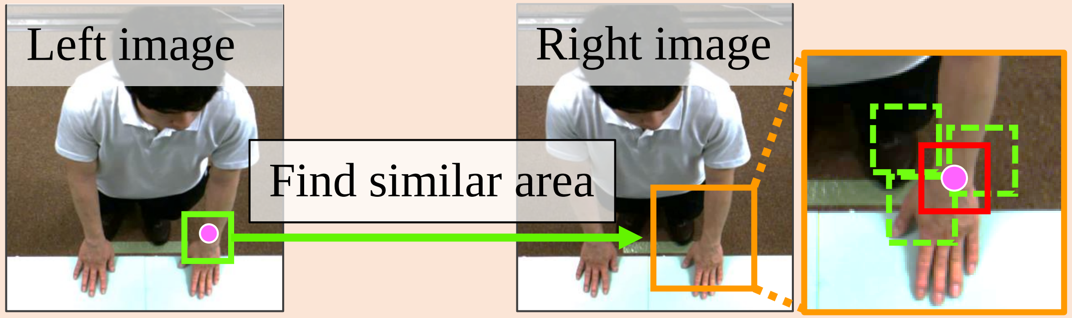
3D human pose estimation
based on triangulation

3D body joint position calibration
using 3D reference points

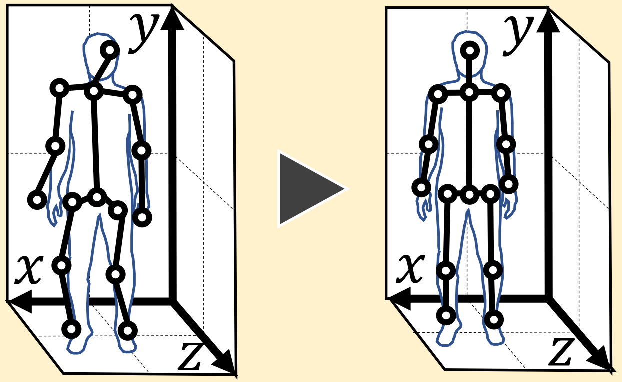
Keep the consistency of the 2D body joint position.



Identify plausible 2D body joint position in the stereo image.



Minimize the 3D position error
by least-squares method.



■ Our framework: **Stereo camera-based 3D human pose estimation**

<u>2D body joint position tracking by optical flow</u>

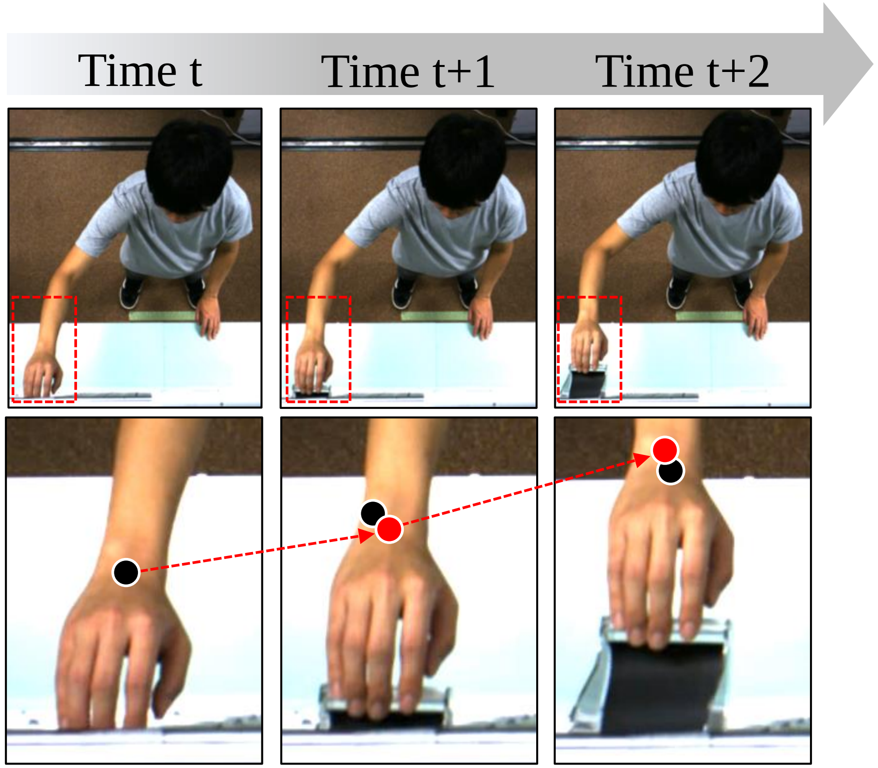
1. Defined as an initial 2D body joint position detected by the CNN at time t.
2. Track the 2D body joint position on the image at time $t+1$ using optical flow.
3. Define the 2D body joint position at time $t+1$ as follow and repeat the tracking at subsequent times.

$$
j_i^{new} = \begin{cases} j_i^{OF} & , \left\| j_i^{OF} - j_i^{CNN} \right\|_2 < 5px \\ mid(j_i^{CNN}, j_i^{OF}) & , \left\| j_i^{OF} - j_i^{CNN} \right\|_2 \geq 5px \text{ and } \\ & \left\| j_i^{OF} - j_i^{CNN} \right\|_2 < 10px \\ j_i^{CNN} & , \left\| j_i^{OF} - j_i^{CNN} \right\|_2 \geq 10px \end{cases}
$$

j_i^{OF} : Tracked 2D body joint position using optical flow

j_i^{CNN} : Detected 2D body joint position by CNN

$mid(j_i^{CNN}, j_i^{OF})$: Calculate the midpoint between 2 points



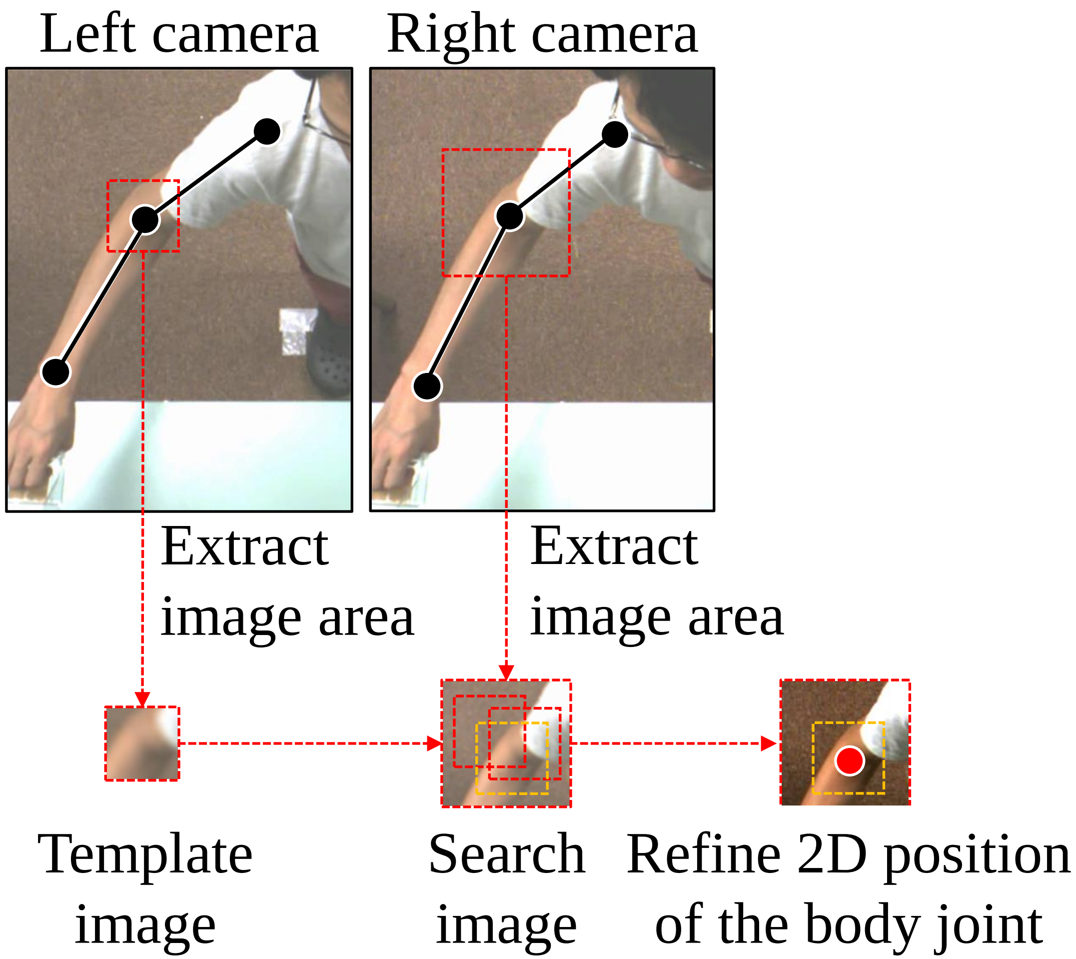
● Detected 2D body joint position by CNN

● Tracked 2D body joint position using optical flow

■ Our framework: **Stereo camera-based 3D human pose estimation**

<u>2D body joint correction using template matching</u>

1. Generate a template image and a search image within a certain range from the 2D body joints position.
2. Search for the image area of highest similarity in the search image to the template image.
3. The center position of the area is defined as the plausible 2D body joint position.



■ Our framework: **Stereo camera-based 3D human pose estimation**

3D calibration using multiple 3D reference points

1. Multiple reference points area placed entirely in the field of view of a stereo camera.
2. Extract the 2D position of these points on each image.
3. Estimate the 3D position of these points by triangulation.
4. Calculate the matrix A that minimize the 3D position measurement error by fitting 5th order polynomial function using least-squares method.

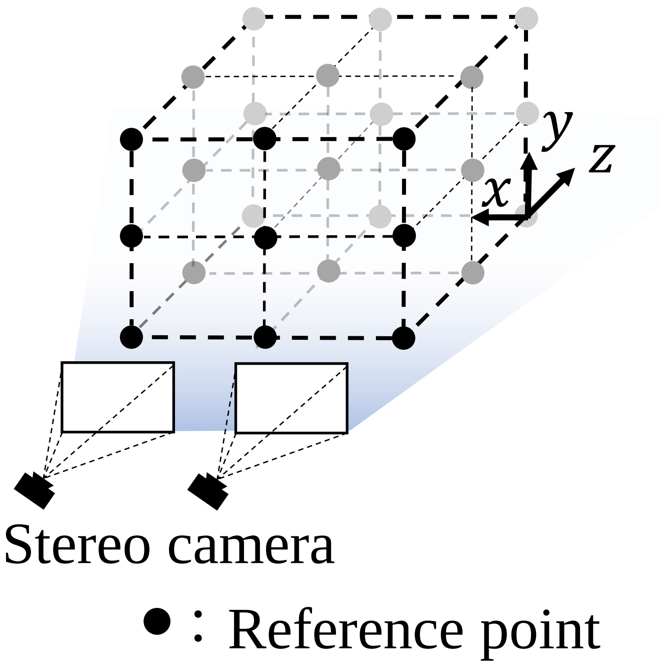
$$\text{argmin} \sum_{i=1}^{n} \| A \cdot P'_i - P_i \|$$

P_i : Actual position of reference points
P'_i : Estimated position of reference points
n : The number of reference points

5. Refine the 3D position of the body joint using the matrix A.

$$J_i^{new} = A \cdot J_i$$

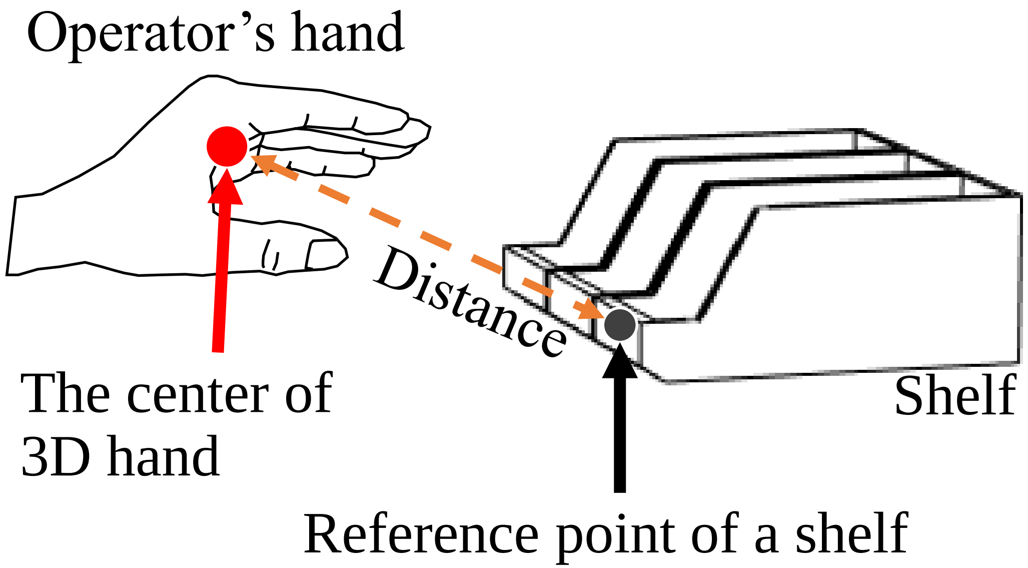
J_i^{new} : Refined 3D body joint position
J_i : Estimated 3D body joint position



■ Our framework: **Picking task determination using 3D hand**

<u>Procedure</u>

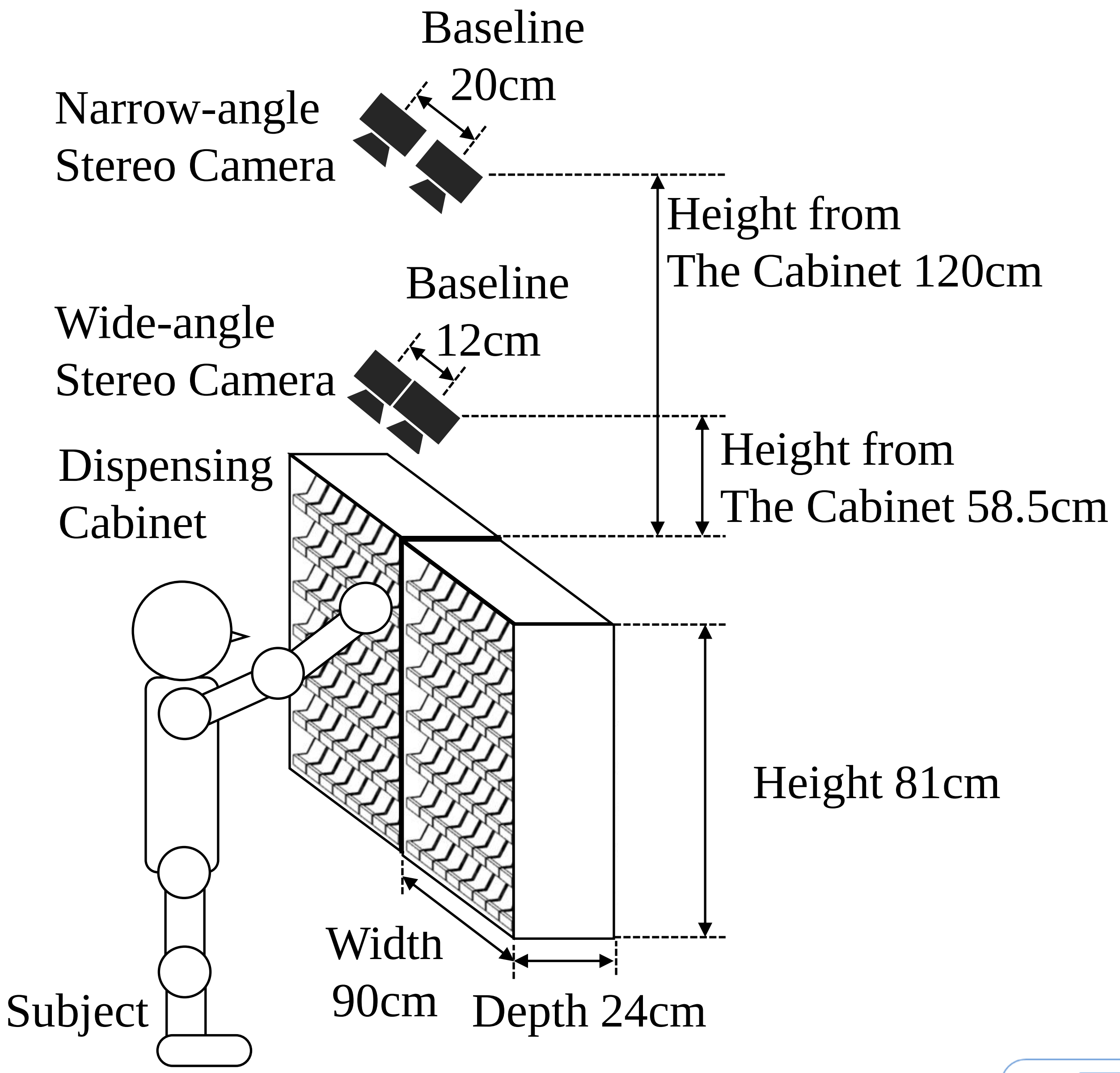
1. Calculate the center position of the 3D hand landmark as the 3D hand position.
2. Identify the closest shelf by calculating the distance between the 3D hand position and shelves.
3. Determine the shelf as “operated shelf” if it is detected as closest for more than 0.5 seconds.



■ Experiments

The experimental setup

- Dispensing cabinet
  - Capacity: 63 shelves (7 rows×9 columns)
  - A shelf size: 9.4cm × 10.6cm × 13.3cm
- Stereo camera
  - Narrow-angle stereo camera
    - Field of view: 48.5° × 36.9°
    - Baseline length: 20cm
  - Wide-angle stereo camera
    - Field of view: 120° × 100°
    - Baseline length: 12cm

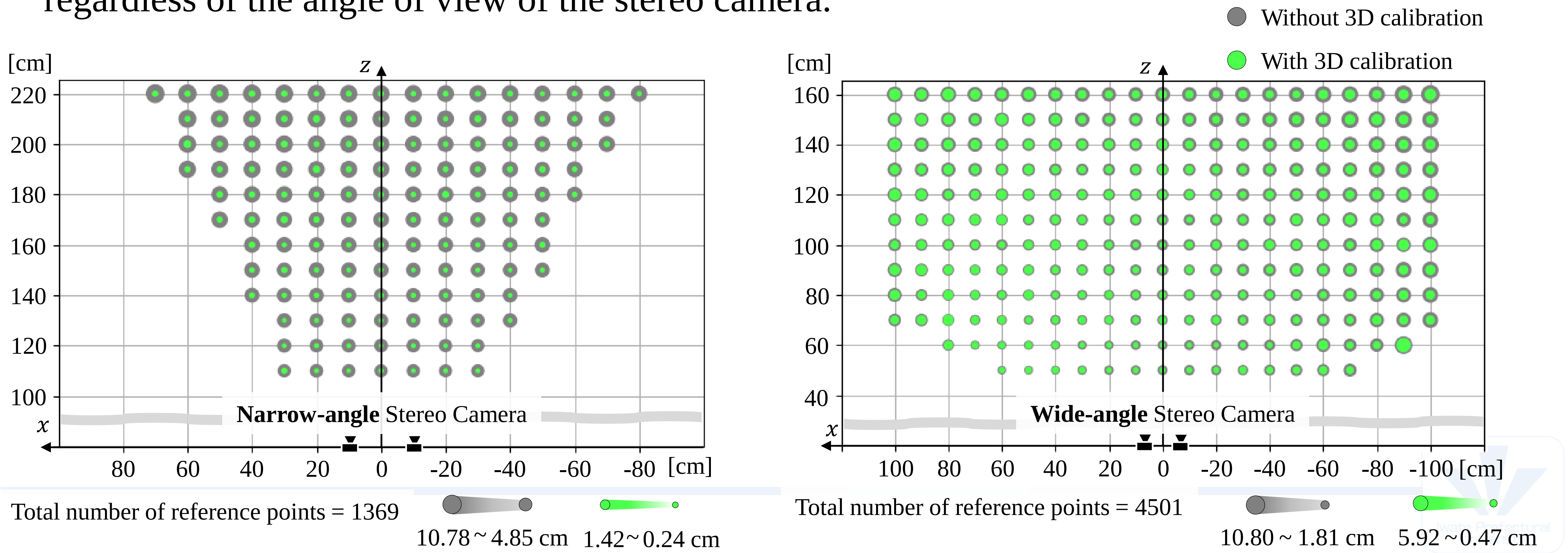


■ Experiment 1: **3D point measurement accuracy by stereo cameras**

- Estimate the 3D position of multiple reference points in the field of view of each stereo camera.
- Calculate Root Mean Square Error (RMSE) of 3D point measurement using each stereo camera.

➢ Result

The 3D calibration method can improve the 3D point measurement accuracy regardless of the angle of view of the stereo camera.



■ Experiment 1: **3D point measurement accuracy by stereo cameras**

- Estimate the 3D position of multiple reference points in the field of view of each stereo camera.
- Calculate Root Mean Square Error (RMSE) of 3D point measurement using each stereo camera.

➢ Result

The 3D calibration method can improve the 3D point measurement accuracy regardless of the angle of view of the stereo camera.

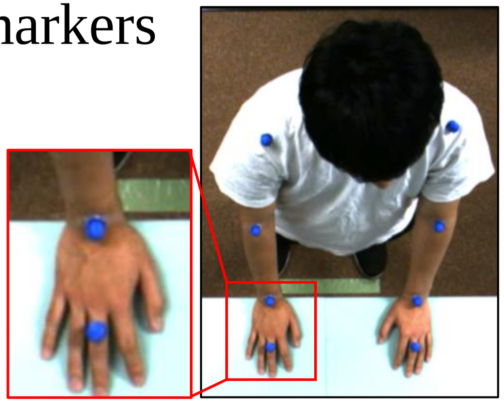
Stereo Camera	Applying 3D Calibration		
	Before RMSE	After RMSE	Δ RMSE
Narrow-angle	7.8cm	0.8cm	-7.0cm
Wide-angle	5.0cm	1.8cm	-3.2cm

■ Experiment 2: **Accuracy of 3D human pose estimation**

- Estimate the 3D position of body joints during the picking task of the subject's attached markers.
- Compare the estimation accuracy of the 3D human pose with and without the 2D body joint correction.

➢ The location of markers

- Shoulder
- Elbow
- Wrist
- Hand



➢ Picking task setting

- Picking task (6 shelves can be measured by <u>both stereo camera</u>)
  → A-6, D-6, G-6, A-13, D-13, G-13
- Wide-area picking task (4 shelves can be measured by <u>only the wide-angle stereo camera</u>)
  → A-1, A-18, G-1, G-18

➢ The location of shelves

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

[grey box] shows 74 shelves can be measured by narrow-angle stereo camera.

[green box] shows shelves picked on the picking task.

[orange box] shows shelves picked on the wide-area picking task.

Iwate Prefectural University

■ Experiment 2: **Accuracy of 3D human pose estimation**

Accuracy {**RMSE (SD)**} of 3D human pose estimation using the narrow-angle or the wide-angle stereo camera during the picking task for shelves which can be measured by both stereo camera [cm]

Body Joint		Methods							
		None		Optical Flow		Template Matching		Optical Flow and Template Matching	
		Narrow	Wide	Narrow	Wide	Narrow	Wide	Narrow	Wide
Right	Shoulder	4.8 (2.1)	8.3 (3.6)	4.8 (2.1) -	7.6 (3.4) ↓	2.5 (1.0) ↓	3.7 (1.3) ↓	2.4 (0.9) ↓	3.5 (1.2) ↓
	Elbow	5.6 (2.7)	6.5 (3.1)	5.5 (2.5) ↓	5.9 (2.8) ↓	2.8 (1.6) ↓	3.6 (1.7) ↓	2.8 (1.6) -	3.6 (1.7) -
	Wrist	4.8 (2.1)	8.1 (4.5)	4.5 (2.0) ↓	8.4 (4.9) ↑	1.6 (0.7) ↓	2.8 (1.3) ↓	1.6 (0.7) -	2.9 (1.3) ↑
	Hand	3.1 (0.9)	3.8 (1.5)	3.0 (0.9) ↓	3.8 (1.5) -	2.1 (0.8) ↓	2.6 (1.1) ↓	2.2 (0.8) ↑	2.6 (1.1) ↑
Left	Shoulder	4.3 (2.1)	7.4 (3.7)	4.2 (2.1) ↓	6.3 (3.3) ↓	2.0 (0.9) ↓	2.5 (1.1) ↓	1.9 (0.8) ↓	2.4 (1.0) ↓
	Elbow	4.9 (2.4)	6.5 (3.6)	5.1 (2.4) ↑	6.2 (3.5) ↓	1.9 (0.7) ↓	2.8 (1.1) ↓	1.8 (0.7) ↓	2.8 (1.1) -
	Wrist	4.4 (2.1)	6.6 (3.4)	4.1 (1.9) ↓	7.0 (3.9) ↑	2.0 (0.8) ↓	2.8 (1.1) ↓	2.0 (0.8) -	2.9 (1.2) ↑
	Hand	3.0 (1.1)	4.2 (1.8)	3.0 (1.1) -	4.2 (1.8) -	2.1 (0.8) ↓	2.6 (0.9) ↓	2.1 (0.8) -	2.6 (0.9) -
Mean of RMSE		4.36	6.43	4.28	6.18	2.13	2.93	2.10	2.91

■ Experiment 2: **Accuracy of 3D human pose estimation**

Accuracy {**RMSE (SD)**} of 3D human pose estimation using the wide-angle stereo camera during the picking task for shelves which can be measured by only the wide-angle stereo camera [cm]

Body Joint		Methods: None	Optical Flow	Template Matching	Optical Flow and Template Matching
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

■ Experiment 2: **Accuracy of 3D human pose estimation**

Results of parametric multiple comparison test for the accuracy of 3D human pose with and without the 2D body joint correction during the picking tasks

Tukey-Kramer Multiple Comparison Test			Experiment Setting		
			Narrow-angle stereo camera	Wide-angle stereo camera	
			Picking Task	Picking Task	Wide-area Picking Task
			P-value	P-value	P-value
None	vs	Optical Flow	.994	.974	.840
None	vs	Template Matching	< .001	< .001	< .001
None	vs	Optical Flow and Template Matching	< .001	< .001	< .001
Optical Flow	vs	Template Matching	< .001	< .001	< .001
Optical Flow	vs	Optical Flow and Template Matching	< .001	< .001	< .001
Template Matching	vs	Optical Flow and Template Matching	.999	.999	.982

■ Experiment 3: **Determination accuracy of the picking task**

Verify the proposed framework determine the correct picking task.

➢ Measurement in different tasks

- Picking task for shelves which can be measured by <u>both stereo camera</u> (**74 shelves**)
- Picking task for shelves which can be measured by <u>only the wide-angle stereo camera</u> (**All shelves**)

Cabinet 1

A-1	A-2	A-3	A-4	A-5	A-6	A-7	A-8	A-9
B-1	B-2	B-3	B-4	B-5	B-6	B-7	B-8	B-9
C-1	C-2	C-3	C-4	C-5	C-6	C-7	C-8	C-9
D-1	D-2	D-3	D-4	D-5	D-6	D-7	D-8	D-9
E-1	E-2	E-3	E-4	E-5	E-6	E-7	E-8	E-9
F-1	F-2	F-3	F-4	F-5	F-6	F-7	F-8	F-9
G-1	G-2	G-3	G-4	G-5	G-6	G-7	G-8	G-9

Cabinet 2

A-10	A-11	A-12	A-13	A-14	A-15	A-16	A-17	A-18
B-10	B-11	B-12	B-13	B-14	B-15	B-16	B-17	B-18
C-10	C-11	C-12	C-13	C-14	C-15	C-16	C-17	C-18
D-10	D-11	D-12	D-13	D-14	D-15	D-16	D-17	D-18
E-10	E-11	E-12	E-13	E-14	E-15	E-16	E-17	E-18
F-10	F-11	F-12	F-13	F-14	F-15	F-16	F-17	F-18
G-10	G-11	G-12	G-13	G-14	G-15	G-16	G-17	G-18

■ shows 74 Shelves can be measured by narrow-angle stereo camera

➢ Result

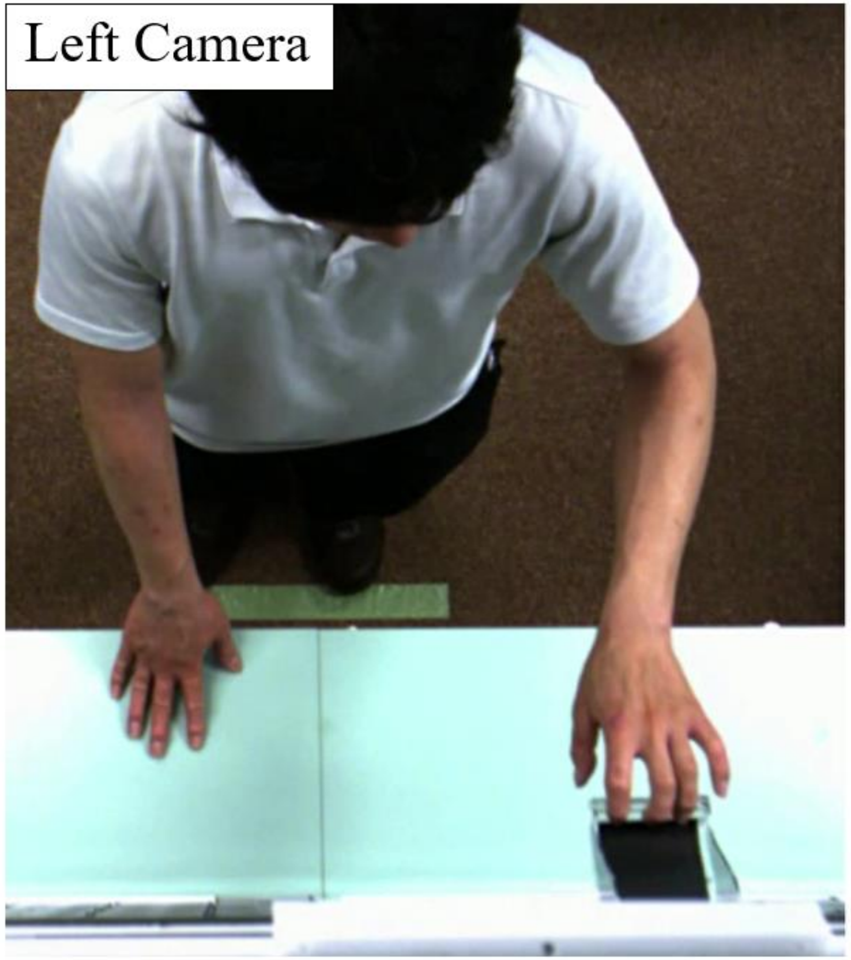
- Enabled to determine the picking task accurately.
- Determination with the wide-angle stereo camera were more accurate.

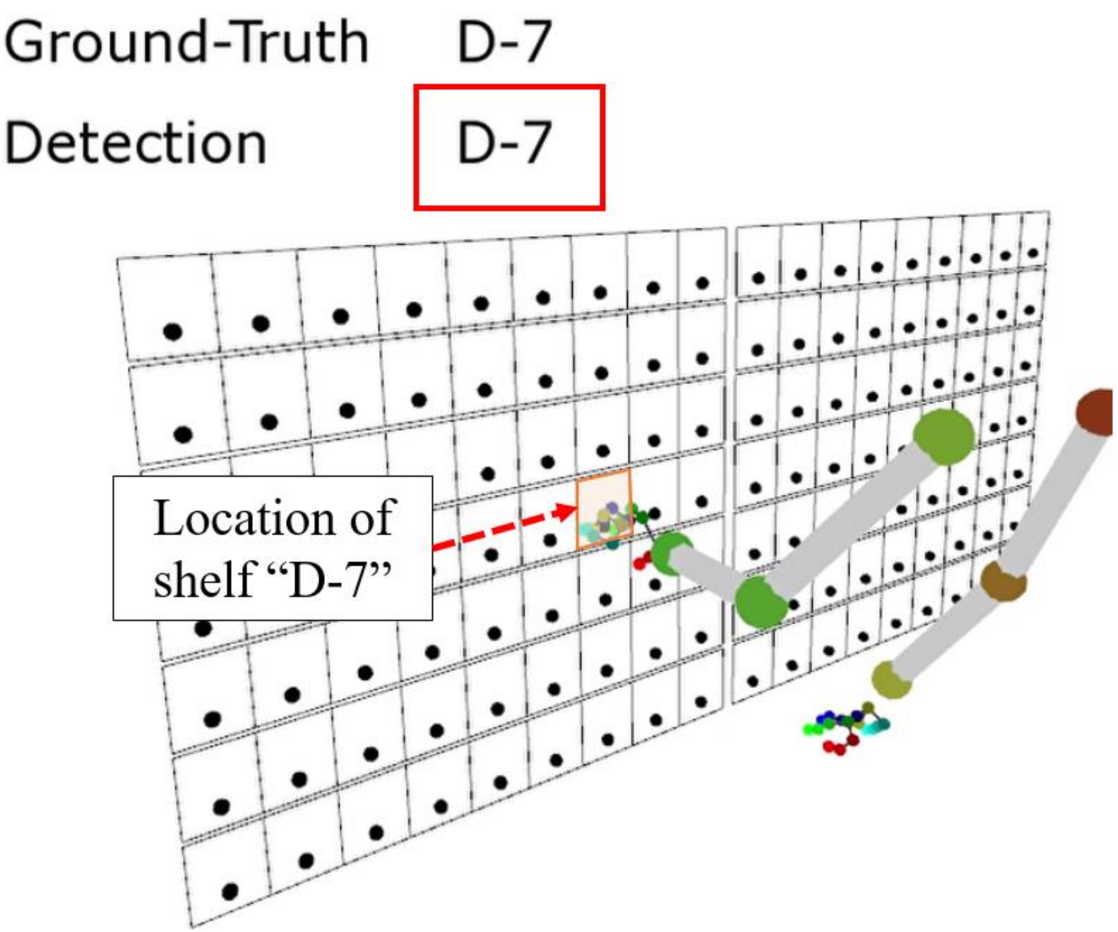
	Stereo Camera		
Subject	Narrow-angle	Wide-angle	
	74 shelves	74 shelves	All shelves
A	91.9%	<u>95.9%</u>	92.9%
B	93.2%	<u>100.0%</u>	93.7%
C	100.0%	<u>100.0%</u>	95.2%
D	95.9%	<u>100.0%</u>	96.0%

Iwate Prefectural University

■ Picking task determination using the narrow-angle stereo camera

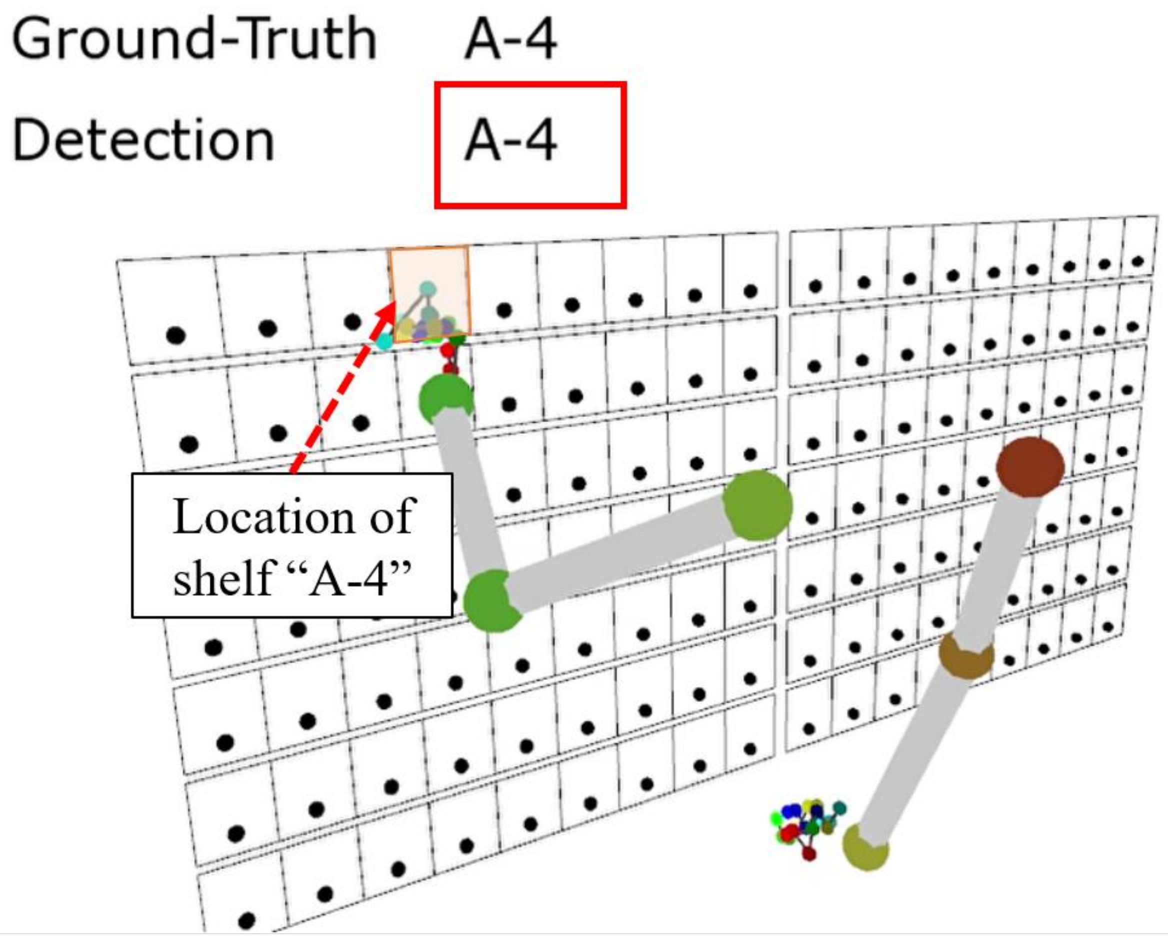
Our framework using the narrow-angle stereo camera can determine the picking task.





■ Picking task determination using the wide-angle stereo camera

Our framework using the wide-angle stereo camera can also determine the picking task.



■ Conclusion

Achievement

- Accurate estimation of 3D position regardless to the view angle of camera by 3D calibration.
- Acceptable accuracy of 3D human pose estimation for determination of picking tasks using template matching.
- Accurate determination of the picking tasks based on the 3D hand of the operator.

Future works

- Improve the determination algorithm by combining the 3D hand position with the 3D position of the operated shelf.
- Verify the evaluation of the physical workload of actual picking task by extending the framework.

■ Reference

[1] ZEBRA Technologies Corp., DS8100-HC SERIES 1D/2D HANDHELD IMAGERS, https://www.zebra.com/us/en/products/scanners/healthcare-scanners/ds8100-hc.html

[2] YUYAMA MANUFACTURING CO., LTD., https://www.yuyama.co.jp/product/products/web_catalog/DrugStation/html5.html#page=3

[3] AOIO SYSTEMS CO., LTD., Projection Picking System (PPS), https://www.hello-aioi.com/en/solution/digital_picking/projection/pps/

[4] 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

[5] 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

[6] Pharmaceutical Safety and Environmental Health Bureau, “The state of dispensing operations”, https://www.mhlw.go.jp/content/000498352.pdf (in Japanese)

[7] 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

[8] 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