Picking and Assortment Operation Assistance Systems with the Depth Camera

Yoshitoshi Murata, Yuki Takehara
Faculty of Software and Information Science
Iwate Prefectural University
Takizawa, Japan
e-mail: y-murata@iwate-pu.ac.jp, g031m104@s.iwate-pu.ac.jp

Yoshihiro Uda
Graduate School of Software and Information Science, Iwate Prefectural University
Takizawa, Japan
e-mail: uda@ieee.org

Takamichi Yoshikawa
Aioi Systems Co., Ltd
Ota-ku, Tokyo, Japan
e-mail: yoshikawa@hello-aioi.com

Abstract*—Industry 4.0 is a well-regarded concept for automation in manufacturing. However, a shortage of high-skilled workers has necessitated realistic solutions for establishing high productivity and quality. We propose an information and communications technology (ICT) picking assistance system to lower human errors for high quality. In this system, a depth camera detects whether a hand is inserted into the correct cell of a shelf to pick items and whether a hand is inserted into the correct box in a cart to put items in. MS-KINECT is used as a depth camera. The misdetection rate for wrong operation in the picking process is very low in an experimental evaluation, and we expect it to be very close to zero in the near future. We determine that the proposed picking and assortment error detection function would be useful for business. In this system, projection mapping technologies are used to indicate which cell items should be picked from, instead of using a lamp. The indicating system, which uses a projector, has a lower introduction cost than those using a lamp. We clarify that gray sandpaper is one of the best materials to serve as a tag for MS-KINECT to recognize indicated colors and digits.

Keywords—Smart factory; Industry 4.0; picking; depth camera; MS-KINECT; projection mapping.

I. INTRODUCTION

We are developing picking assistance system [1]. The German government's Industry 4.0 (ID4) initiative *Industrie 4.0* has revolutionized Germany's manufacturing industry [2], with goods in "smart factories" being moved, picked, and delivered automatically [3]. ID4 technologies are useful for establishing high productivity and quality in light of a shortage of high-skilled workers. If the latest manufacturing robotics and custom assembly lines were introduced in model factories for ID4, products would be automatically conveyed and assembled, and there would be fewer workers. However, it is impossible for most existing factories to replace all of their manufacturing lines with more advanced ones. Realistic solutions for establishing high productivity and quality in light of a shortage of high-skilled workers are as follows:

(1) Developing and introducing information and communications technology (ICT) systems to bring low-skilled workers closer to the level of high-skilled workers and to lower human errors.
(2) Replacing workers with robotics step by step.

As in manufacturing, picking processes have been widely introduced in logistic centers. Several kinds of robot warehouses are being developed and introduced [4–6]. These robot warehouses save workers from having to navigate large warehouses to accomplish picking. Bins or shelves in which goods are stored come to work stations for workers or robots to place the goods into a box. The goods robots can pick are limited, such as stationaries or packaged goods. Robots have difficulty picking food materials or diversely shaped goods. Therefore, workers are still required in the picking process.

In this paper, we propose an ICT picking and assortment assistance system to lower human errors in the picking process. The picking process is when a worker picks items, such as assembly parts out of numbered cells on shelves and puts them into assortment boxes. For example, in an automobile assembly factory, a worker takes different parts from cells of a shelf and puts them into assortment boxes corresponding to production orders. The parts in each cell are the same, and different items are stored in different cells. Assortment boxes are carried to workers on an assembly line. If a worker picks parts from an incorrect cell and the parts are subsequently assembled into a product, it would take too much time to detect the incorrect parts and exchange them with correct ones. In the worst case, an automobile assembled with incorrect parts could be shipped.

Even though picking operations are monotonous, completely preventing mistakes is difficult because workers are human. Therefore, there are several kinds of picking assistance systems for decreasing incorrect pickings. Aioi Systems Co. Ltd., to which one of the co-authors belongs, provides the digital picking system "L-PICK," which indicates the cells of a shelf and the number of items to pick from those cells by lighting a lamp mounted on each cell [7]. However, since L-PICK does not have a function that detects incorrect picking, completely preventing it is impossible. Many companies request Aioi Systems Co. Ltd. to develop and provide the picking error detection system to lower recovery cost. Hence, we developed an operation error detection system for the picking process with MS-KINECT [8] as the depth camera in first version [9]. In this system, two sets of MS-KINECT traced a hand wearing a colored glove from diagonally backward and monitor whether the hand was inserted into a cell. Since this system used a color tracing function to measure the position of a hand, it could not be applied to a food material delivery service because of differences in food color. Therefore, we proposed the new picking assistance system as the next version in ACHI 2018 [1]. In case of the picking assistance system, a set of MS-KINECT was mounted on the top of a shelf. Since our partner Aioi Systems Co. Ltd. developed a new cell lighting technology that used a projector [7], we used it in our new system. The system recognized the position of a cell lit by a projector and the number presented on a tag, detected the position of a cell into which a hand was inserted, counted the number of times a hand had been inserted into the cell, and compares them with recognized ones.

In addition to the picking assistance system, we also developed an assortment assistance system. In this system, MS-KINECT was mounted above an assortment of boxes to detect whether a hand that had items was inserted into the correct assortment box. The misdetection rate for wrong operations during picking was low in an experimental evaluation, and we expect it to be close to zero in the near future. However, the assortment detection accuracy was determined to be unacceptable in the experimental evaluation.

In this paper, we propose not only a picking assistance system but also a new assortment assistance system for boxes, in which the misdetection rate for wrong operation is low.

After introducing related works in Section II, detection methods for incorrect operations are introduced in Section III. Detection technologies in which MS-KINECT is used are introduced in Section IV. Experiments and results are described in Section V. Conclusions and future work are described in Section VI.

II. RELATED WORK

Human beings have excellent abilities. Workers in an automobile assembly factory can use their sense of vision and touch to detect subtle depressions or distortions that a computer system cannot. On the other hand, human beings sometimes make mistakes. Several kinds of assistance systems that decrease the number of mistakes have therefore been developed. Existing picking assistance systems are introduced in this section.

A picking assistance system has some of the following four functions:

(1) Indicating a cell of a shelf from which an item should be picked and the number of items that should be picked from that cell.
(2) Detecting whether a worker has picked the correct number of items from the correct cell.
(3) Indicating the box or cell of a tray in which picked items should be stored.
(4) Detecting whether a worker has put items in the correct box or cell of a tray.

There are several kinds of system for indicating cells. Aioi Systems Co. Ltd. provides the digital picking system "L-PICK," which indicates cells of a shelf and the number of items to pick from those cells by lighting a lamp mounted on each cell [3]. There are several systems in which a Head-Up Display (HUD) and augmented reality technology are used to assist picking operations. Schwerdtfeger used a semi-transparent HUD and augmented reality technology enabling a worker to see an arrow or frame displayed in front of a cell of a shelf [10]. Baumann used a single-eye HUD, and a worker recognized a cell from which he or she would pick up items with guidance displayed on a mirror of the HUD [11][12]. Guo compared the HUD-based systems with the cart-mounted display (CMD), Light, and Paper Pick List as picking assistance systems [13]. This system also provided the delivery assistance function. They concluded that the pick-by-HUD and pick-by-CMD were superior by all metrics than the current pick-by-paper and pick-by-light systems, but the differences between the HUD and CMD were not significant and did not show that a HUD was better than a CMD. However, experimental results should be different in other experimental conditions. In practical situations, the number of cells (in that paper they were called bins) is usually less than 12, and a worker can see a lighted lamp at a glance. And multiple lamps are not lighted simultaneously; a single lamp is lighted for each occurrence where an item is picked up. Therefore, the practical error rates and task times recorded in this study would produce better results than those in that study. Furthermore, it is not certain whether a worker should have a palm-size PC and wear a HUD for extended periods. In our research, most workers would not like to carry a barcode reader.

As described above, systems indicating a cell by a lamp have been used in business. Likewise, indicating systems using a HUD have been developed in research.

We consider a picking and assortment error detection method in the next section.

III. PICKING AND ASSORTMENT ERROR DETECTION METHODS

In this section, prospective methods for detecting whether items are picked from the correct cell and assorted to the correct box are introduced and evaluated. This time, in addition to detecting when items are picked from a cell, these methods determine when a hand is inserted into a cell, when a tag, such as the barcode attached to a cell is read, and when a picked item is dropped, and a new item is picked. The following are the prospective methods:

(1) Reading a **barcode** attached on a cell with a barcode reader.
(2) Reading a **passive Radio Frequency (RF)-ID** set on a cell with a RF-ID reader.
(3) Reading an **active RF-ID** set on a cell with a RF-ID reader.
(4) Detecting change of weight with a **load sensor**.
(5) Detecting when a hand and/or arm is inserted into the correct cell with a **photoelectric sensor**.
(6) Detecting when a hand and/or arm is inserted into the correct cell with a **depth camera,** such as MS-KINECT.

The above methods are narrowed down by the following evaluation criteria:

(1) Additional cost to introduce a detection function.
(2) Additional operations for a worker.
(3) Detection accuracy.

An evaluation of the picking error detection methods is shown in Table I. As for barcodes, the cost of attaching a barcode to each cell is cheap, and barcode readers are not expensive. However, carrying a barcode reader and scanning barcodes are cumbersome for workers.

Passive RF-ID presents the same difficulties as barcodes. In addition to having to carry a RF-ID reader, the weak signal strength of passive RF-ID requires positioning the reader in close proximity to a RF-ID tag.

As for active RF-ID, despite having to carry the reader, it does not need to be positioned in close proximity to a RF-ID tag because the signal strength is strong. However, because of their strong signal strength, active RF-ID readers sometimes read RF-ID tags placed in other cells.

As for load sensors, their detection accuracy is high. However, they are usually expensive, and each sensor must be wired to a PC. Introduction costs are therefore high. The same holds true for photoelectric sensors.

A depth camera using MS-KINECT usually cost a few hundred dollars. While introduction costs would be high under our proposed system because one MS-KINECT set would be required per shelf, our system alleviates the need for workers to carry a reader, and the detection accuracy is high. We have determined that the depth camera would be the best method overall for our picking assistance system.

The following three prospective methods are considered for delivery:

(1) Detecting change of weight with a **load sensor**.
(2) Detecting whether a hand is inserted into the correct box with a **photoelectric sensor**.
(3) Detecting whether a hand is inserted into the correct box with a **depth camera,** such as MS-KINECT.

The evaluation criteria for delivery is the same as those for picking. The evaluations for the above three methods are the same as those in Table I. We think a method using a depth camera camera is the best for delivery when its accuracy is high.

TABLE I. EVALUATION OF PICKING ERROR DETECTION METHODS

Method	A. Cost	A. Operation	Accuracy
Barcode	Low	Big	Middle
Passive RF-ID	Middle	Big	Low
Active RF-ID	Middle	Little	Middle
Load	High	Little	High
Photoelectric	High	Little	Middle
3D camera	Low	Little	This paper

IV. WRONG OPERATION DETECTION TECHNOLOGY IN PICKING

In this section, wrong operation detection technology in which a worker picks an item from an incorrect cell is introduced. Here, the incorrect cell is not the indicated cell from where an item should be picked.

A. Detection algorithm for wrong operation

As described in the previous section, a technology that detects whether a worker picks items from an indicated cell is in demand. Because detecting whether a worker is picking items from a cell is difficult, we decided to focus on detecting whether a hand is inserted into a cell instead of detecting whether a hand picks an item from a cell.

The MS-KINECT is widely used to estimate the motion of the human skeleton. In addition to its skeleton estimation function, it has several useful functions for detecting whether a hand is inserted into a cell such that the positions of joints on the body, the position of an indicated point, and the depth of a position can be measured, and the edges of the body can be recognized. And, its introduction cost is reasonable. Hence, we decided to use MS-KINECT to detect whether a hand is inserted into an indicated cell. The following schemes were considered as alternatives:

(1) Skeleton scheme: position of the hand's joint in the skeleton is used.
(2) Body edge scheme: position of a recognized hand edge is used.
(3) Color tracing scheme: position of a hand wearing a colored glove is used.
(4) Depth change scheme: position of a change in depth is used.

Because the skeleton estimation function in MS-KINECT needs video of a hand, an MS-KINECT must be placed from 0.5–5 m in front of a body. And, the accuracy of an estimated position is best when it is placed right in front of the body. Its accuracy worsens when it is placed more diagonally. Because a worker stands in front of a shelf, this scheme is not suitable for picking operations. As with the skeleton scheme, placing the MS-KINECT in a suitable position for detecting the edge of a hand is difficult.

Therefore, we experimentally evaluated the color tracing scheme and depth change scheme as follows.

B. Color tracing scheme

Because setting an MS-KINECT in front of a worker in a factory or delivery center is impossible, we set the MS-KINECT behind the workers. In this setting, because it is impossible for a single camera to always watch a hand that is hidden by the placement of its body, we placed two MS-KINECTs on both sides of a worker as shown in Fig. 1. And, because the skeleton estimation function and body edge recognition function are not used to trace a hand, the Open-CV color tracing technologies [14] are used to trace hands. From experience, red, orange, yellow, and yellow-green are used for color tracing.

The coordinates of the four corners of each cell are pre-set before estimating the cell number. Because we use a shelf as shown in Fig. 2 in the following experiments, the coordinates of 16 corners are pre-set as in Fig. 3.

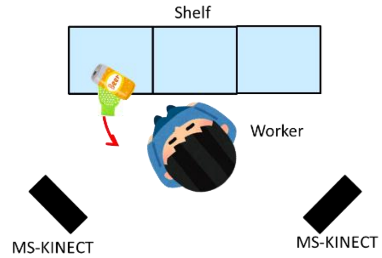

Figure 1. Layout of a worker, shelf, and MS-KINECTs (top view)

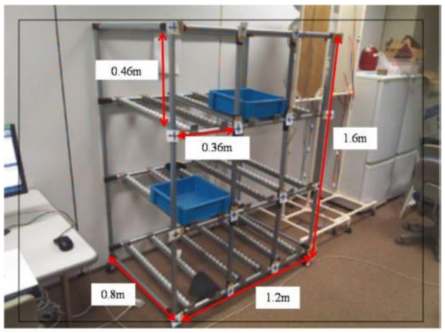

Figure 2. Shelf used in experiment

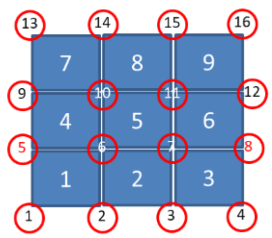

Figure 3. Pre-set coordinates for corners of cells in shelf

The error detection system that uses this scheme may count picking operations multiple times in one operation because of graphical noise or worker motions. The following procedures were used to decrease the multiple counts;

(1) Recognition times of a decided color pixel: Because the position of graphical noise artifacts changes randomly, we recognized that the decided color should be detected multiple times at the same position. In the following experiment, five times was used as the number of determination times.

(2) Recognition guard time: Because most workers could not quickly pick an item from a cell, we set a guard time in which the system does not count formerly recognized operations again. In the following experiment, we use 0.25 sec as the guard time.

(3) Recognition position of re-detected glove: The system loses sight of a glove when the glove is hidden by a worker's body. Because the glove's position is usually in the same position as when the glove is lost, the system does not count the glove's motion as a picking operation.

We performed two experiments. The shelf shown in Fig. 2 was used in these experiments and had nine cells. In the first experiment, participants picked items from each cell ten times without any operation guidance. In the second one, participants performed operations the same as they did in the first experiment after receiving operation guidance. The guidance cautioned participants about picking motions that caused misrecognition. The practical cautions are described later. The number of participants in the first experiment was six, and for the second experiment, five. Some participated in both experiments, and others in only one of them.

We independently counted the number of times that each MS-KINECT recognized a worker picking an item from a correct cell (a correct operation is detected) and the number of times each MS-KINECT recognized a worker picking an item from an incorrect cell (a correct operation is interpreted as being incorrect). From these two data sets, we calculated that either of the two MS-KINECTs detected a worker picking from a correct cell or from an incorrect cell. Beer cans were used as picking items.

The results of the first experiment are shown in Table II. The number of correct detections and the erroneous recognition rate are averaged for nine cells. Because the accuracy of recognition in the proposed system may be affected by the height of the MS-KINECTs, we placed the MS-KINECTs at 1.6 m and 1.8 m.

Unfortunately, the proposed system sometimes counted one operation as two operations or recognized an incorrect operation. Since the participants B, C, and D moved their bodies naturally, the number of correct detections was low, and the misdetection rate was also low. However, because participant A did not move in accordance with the position of a cell, his posture was unnatural. The system lost sight of his hand, which was hidden by his body. Participant E moved his hand close to the aperture of a shelf. Therefore, the system sometimes counted one operation twice and mistook the cell from which an item was picked. Because participant F operated slowly, the system sometimes counted one operation twice.

Based on our reflections in the first experiment, we cautioned the participants to perform the following operations in the second experiment:

(1) Move the body in front of a designated cell prior to picking an item.

(2) Do not pick an item from a cell extremely slowly (items should be picked at normal speed).

(3) Pick an item vertically from a cell or move a hand along the aperture of a shelf.

TABLE II. RESULTS OF FIRST EXPERIMENT (WITHOUT GUIDANCE)

Participant	Height of participant (m)	Hight of MS-KINECT			
		1.6 m		1.8 m	
		Number of cerrect detection	Erroneous recognition rate (%)	Number of cerrect detection	Erroneous recognition rate (%)
A	1.74	9.2	0	8.8	0.44
B	1.7	10	0	9.9	0.33
C	1.62	9.8	0.11	9.8	0.11
D	1.75	10	0	9.9	0.11
E	1.66	9.6	0.22	9.7	1.78
F	1.69	10.4	0	10	0.44
Average	1.71	9.83	0.055	9.68	0.54

The results of the second experiment are shown in Table III. Because participant G forgot about the above cautions halfway through experiment, his experimental data were not good. However, the experimental data of other members were good. And, data gathered from the MS-KINECT at 1.6 m were better than those gathered at 1.8 m.

TABLE III. RESULTS OF THE SECOND EXPERIMENT (WITH A GUIDANCE)

Participant	Height of participant (m)	Hight of MS-KINECT			
		1.6 m		1.8 m	
		Number of correct detection	Erroneous recognition rate (%)	Number of correct detection	Erroneous recognition rate (%)
A	1.74	10	0	10	0.11
F	1.7	9.9	0	10.1	0.33
G	1.62	9.9	0.11	9.6	0.44
H	1.75	10	0	9.9	0
I	1.66	10.1	0	10	0
Average	1.69	9.98	0.022	9.92	0.176
Average exclude G	1.71	10	0	10	0.11

C. Depth change scheme

The detection rate of correct operation in the color tracing scheme is good enough, and the erroneous recognition rate is low. However, the system needs two MS-KINECTs for each shelf. Because the color tracing scheme is used, applying the scheme to picking colored items such as vegetables is difficult. Hence, we decided to measure the change of depth when a hand is being inserted into a cell.

From past experience, we have determined that the best mount position for an MS-KINECT to detect whether a hand enters a cell is just above the surface of a shelf aperture. The MS-KINECT 3D camera searches for a hand and arm just over the surface of a shelf aperture as shown in Fig. 4. The MS-KINECT must be set at a position in which its 3D camera can observe the entire shelf aperture. This system detects whether a hand is inserted by changing the depth in front of a cell. When a hand and/or arm is inserted into a cell, the depth in such a view is changed from L_f to L_h. A change in depth L_h corresponds to the length between the MS-KINECT and the hand and/or arm. Its position is within the cell aperture in which the hand and/or arm is inserted.

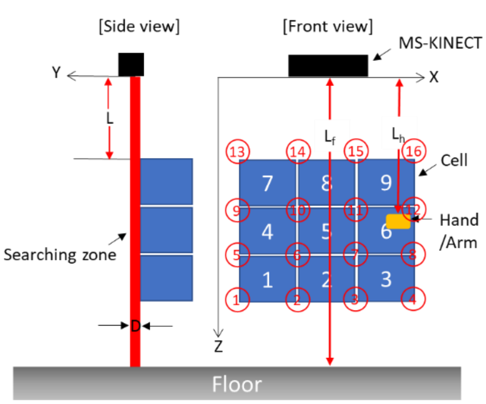

Figure 4. Mounting position of MS-KINECT and searching zone

The coordinates of the four corners of each cell are pre-set before estimating the cell number. In Fig. 4, the coordinates of 16 corners are pre-set. The number of the cell in which a hand is inserted is estimated by comparison between a coordinate of the detected hand and the coordinates of four corners for each cell ⓝ.

We developed an experimental shelf as shown in Fig. 5. An MS-KINECT is mounted 65.5 cm away from the shelf. The shelf consists of 3 x 3 cells. The size of the shelf is 67.5 x 64.5 cm, and the size of each cell is 22.5 x 21.5 cm. The length between the floor and the bottom of the shelf is 98 cm. The reason a shelf in this experiment is different from the shelf showed in Fig. 2 is that it was returned to its owner factory. We measured the error rates for detecting a hand inserted into a cell and whether the MS-KINECT can recognize a lit tag and the number on it using the experimental shelf.

Before estimation, the coordinates of the corners of each cell are pre-set using the pre-set windows shown in Fig. 6. The corner number is selected with the corner number button. The coordinates of each corner are entered by clicking a corner or inputting digits. The red grid of the shelf aperture is generated by clicking the grid button.

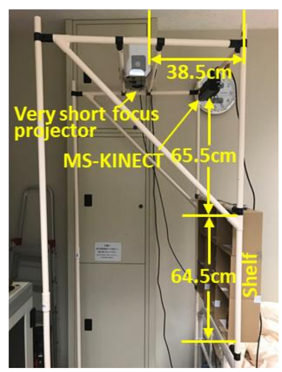

Figure 5. Experimental shelf

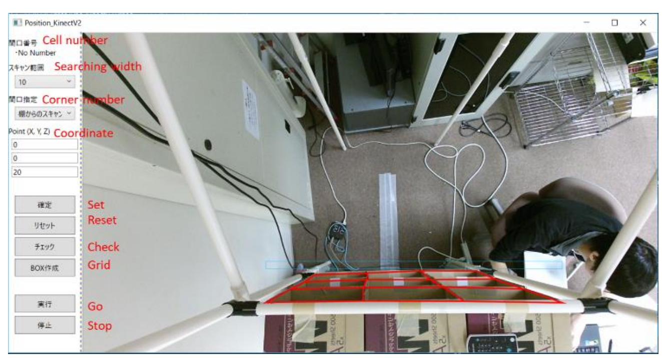

Figure 6. Pre-set window

The erroneous recognition rate for picking operation would change depending on the width D of the searching zone, the searching period, the cell position, and the threshold width to detect a hand and/or arm. We measured the erroneous recognition rate under conditions in which the width D of the searching zone is 1 cm / 3cm / 5 cm / 10 cm, the searching period is 500 msec. / 1000 msec., and the threshold width for detecting a hand and/or arm is 1 cm. Beer cans were used as picking items. The number of participants was ten. Each participant picked an item from each cell five times. The average detection error rates of every cell vs. the width of searching zone D are 1 cm / 3cm / 5 cm / 10 cm as shown in Fig. 7. The parameter of this figure is the searching period. The detection error rate in 500 msec. is lower than that in 1000 msec. in each the searching width D. The average error rate for each cell in which the searching period is 500 msec. is shown in Table IV. The erroneous recognition rate increases in accordance with an increase in width D. This is because a participant tends to insert his/her hand into a cell through the searching zone in front of other cells.

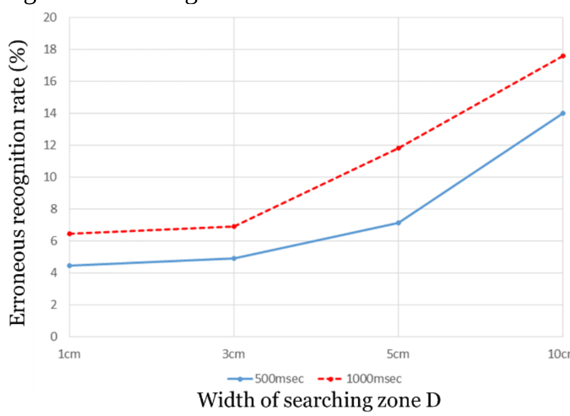

Figure 7. Average erroneous recognition rates for every cell

TABLE IV. AVERAGE ERRONEOUS RECOGNITION RATES (%) FOR EACH CELL

Cell #	1cm	3cm	5cm	10cm
1	2	0	0	4
2	4	4	4	6
3	4	6	4	0
4	4	0	0	1
5	6	6	2	8
6	2	8	14	20
7	6	0	10	24
8	4	16	10	28
9	8	4	20	26

The most serious problem in the picking error detection system is that the system recognizes wrong operations to be fair. We measured the misdetection rate for wrong operation. In this experiment, a correct cell is No. 5. The number of participants is ten. Each participant picks an item from cells around the No. 5 cell five times. The detection rate for wrong operations to be wrong, detection rate for wrong operation to be fair and practical detected wrong operations are shown in Table V. Average detection rate for wrong operations to be wrong is 95.7 %. And, the proposed system did not recognize wrong operations to be fair, completely detected wrong operations.

TABLE V. MISDETECTION RATE FOR WRONG OPERATIONS

Cell #	Detection Rate (%)	No. 5 cell D. Rate (%)	Detected errors
1	98	0	1-6
2	98	0	2-3
3	96	0	3-6, 6
4	98	0	1-4
5			
6	96	0	6-9, 3-6
7	94	0	(4-7) x 3
8	96	0	(5-8) x 2
9	90	0	(6-9) x 5

However, the system recognized that a participant picked an item from the No. 6 cell, even though he picked it from the No. 3 cell. The reason of this error detection is that the system detects an item in front and within 1 cm from the No. 6 cell after picking from the No. 3 cell.

D. Discussion

The depth change scheme is superior to the color tracing scheme concerning the number of MS-KINECT devices and no-limitation for color of goods. However, the misdetection rate of the depth change scheme is higher than that of the color-tracing scheme, as mentioned in Section IV-B. This may be due to the insufficient monitoring accuracy of operation in the depth change scheme. We plan to constitute a few cm non-detection area around each cell, and guard time not to detect after detecting an item to be picked from a cell. We think these constitutions would lower the detection rate for wrong operation to be fair exceedingly close to zero.

V. WRONG OPERATION DETECTION TECHNOLOGY IN DELIVERING

We imagine a delivery cart as shown in Fig. 8. An MS-KINECT is mounted to search the surface of an assortment of boxes and to detect when a worker puts an item into an incorrect box. In this section, the detection technology for wrong operations is introduced.

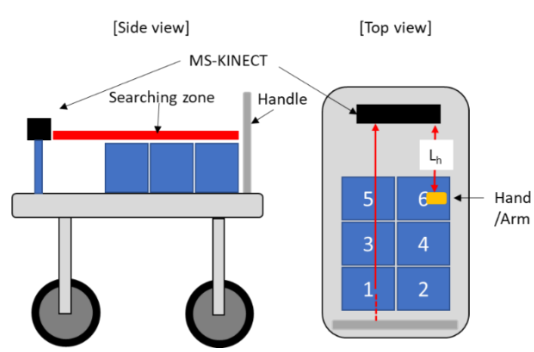

Figure 8. Image of delivery cart

We consider the best position to which the MS-KINECT is mounted to be the front part of the cart. When the MS-KINECT is mounted on the side of the cart, and its position is between an assortment of boxes and a shelf, a worker must pick an item from a cell passing over the cart. We consider this operation to be a bad operation. Hence, we selected the front position for the MS-KINECT.

A. Detection algorithm for wrong operation

We decided to use the MS-KINECT to detect whether a hand is inserted into an indicated cell for item placement. The following schemes were considered as alternatives:

(1) Skeleton scheme: position of the hand's joint in the skeleton is used.
(2) Body edge scheme: position of a recognized hand edge is used.
(3) Color tracing scheme: position of a hand wearing a colored glove is used.
(4) Depth change scheme: position of a change in depth is used.

Because the position to which the MS-KINECT is mounted would be the side of the cart in the first two alternatives, these two schemes are not suitable. And, the skeleton estimation application sometimes mistakenly estimates the position of a hand when the hand is hidden because of its insertion into a box. The reason is that the application tries to find a hand as shown in Fig. 9, even though the hand is hidden by a box. In this figure, the application recognizes a part of the box's partition as a left hand. As for body edge estimation, it sometimes recognizes boxes as being part of a body, as shown in Fig. 10. These are also reasons why we did not select these schemes. Because the third scheme, color tracing, is not suited to colored items such as vegetables, this one also tends to detect wrong operations.

Figure 9. Example of skeleton estimation when a hand is hidden by a box

Figure 10. Example of body edge estimation

Hence, we selected the fourth scheme using a change in depth to detect wrong operations. The MS-KINECT searched just over the surface of the apertures of the boxes as it did in the picking operation in the first experimental system. The maximum depth was far from a screen that was set near a handle. Unfortunately, it is not accurate enough for commercial usage. Therefore, we changed the target area from the screen to the side wall of each box that faces MS-KINECT.

B. The first system using the depth change scheme

The search zone is just above the mass of boxes. When a hand/arm or item is inserted into a box, the PC on this cart detects whether a hand/arm or item is inserted into the box by the change in depth.

The number of the box that a hand is inserted into is estimated to compare it with a coordinate of the detected hand and the coordinates of four corners for each box, the same as for a cell in Section IV-C.

We measured the delivery error rate using six boxes on a table as shown in Fig. 11 instead of using a delivery cart. The MS-KINECT is placed on another table. It was placed 60 cm from the top of the boxes. From the results of experiment in Section IV-C, we decided that the searching period is 500 msec. and that the width of the search zone is 1 cm.

Since we noticed that our system easily detected multiple boxes, we constituted a 5 cm wide non-detection area on boxes that are on the near side of a worker, as shown in Fig. 11. When the depth far from the boxes was not fixed, the detection accuracy was poor and unstable. We set the screen near the box1 and 2 to fix the maximum depth from the MS-KINECT. The average error rate for each box is shown in Table VI. The number of participants was five, the same as in the picking experiment. Each participant puts an item into each box five times. Overall, the error rate, especially the double count rate, is high.

Figure 11. Experimental system for measuring delivery error rate

TABLE VI. AVERAGE DELIVERY ERROR RATES (%) FOR EACH BOX

Box number	No-detection	Double count	Total
1	0	12	12
2	0	8	8
3	8	20	28
4	4	0	4
5	24	4	28
6	0	12	12

C. The second system using the depth change scheme

Because of the accuracy of detecting wrong operations in the above scheme, we developed the other detection technology. In practical terms, we changed the target area from the screen to the side wall of each box that faces the MS-KINECT. The practical layout between the MS-KINECT and assortment boxes and the target areas for detecting the depth change are shown in Fig. 12. Red bars in Fig. 12 are the detection target areas, which are the far edges of each box aperture. Three dimensions of data (x, y, z) from the MS-KINECT of these areas are pre-measured at pixel units and are pre-stored. When a hand hides these areas from the MS-KINECT, the Z values of hidden areas become shorter than the original values. In the case of Fig. 12, the Z values of the 3rd to 6th boxes are changed in accordance with a hand entering the 3rd box to place an item in it. Therefore, we introduced the box decision algorithm as shown in Fig. 13. In this figure, a worker stands to the right side of an MS-KINECT and puts his/her hand into the box from the right side. The relationship between the box into which a hand is put and the boxes of which the Z values at the target area become shorter are as follows:

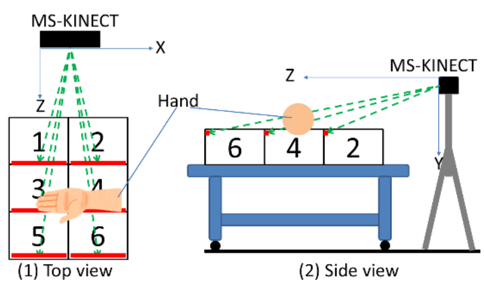

Figure 12. Layout and detecting target areas

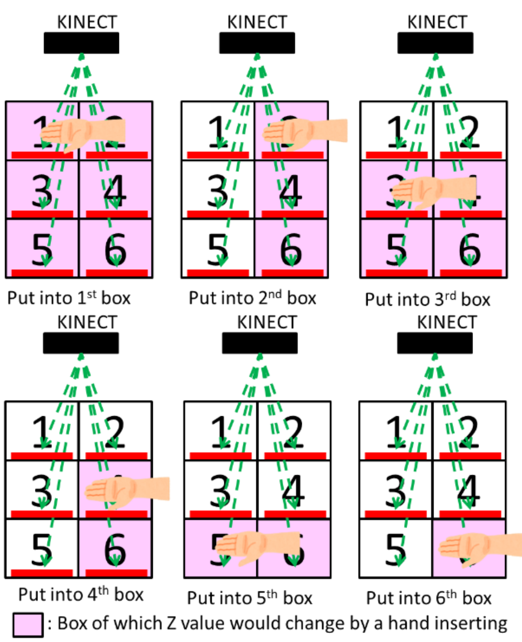

Figure 13. Box decision algorithm

(1) 1st box: Every box.
(2) 2nd box: 2nd, 4th, and 6th box.
(3) 3rd box: 3rd, 4th, 5th, and 6th box.
(4) 4th box: 4th and 6th box.
(5) 5th box: 5th and 6th box.
(6) 6th box: 6th box.

When a worker is standing to the right side of MS-KINECT, the relationship between the box into which a hand is put and the boxes of which the Z values at the target area become shorter are changed. This wrong operation detection scheme must detect a direction that a hand enters from and change the box detection algorithm.

We implemented the application shown in Fig.14 that counts the number of times an item is put into an indicated box. The experimental equipment is shown in Fig. 15. We used copy paper boxes as the assortment of boxes. In the pre-experiment, we noticed that MS-KINECT could watch every aperture and that its height was as low as possible. The practical layout between the MS-KINECT and assortment boxes is shown in Fig. 16.

The number of participants is twelve, and they put a beer can into each box five times. Experimental results are shown in Table VII. The experimental application perfectly detects every participant except participant H. The application made one misrecognition for participant H, recognizing that participant H put an item into the 4th box, not the 2nd box. Participant H was left-handed and put an item into the 2nd box with his left hand. After taking his hand out of the 2nd box, his hand moved over the 4th box, causing the application to make the misrecognition. This misrecognition would be solved by changing the direction of the cart for him so as to enable him to move his hand smoothly, making this scheme for recognizing wrong operation in delivery useful for commercial usage.

We think this depth change scheme is effective to improve the misdetection rate in the picking operation.

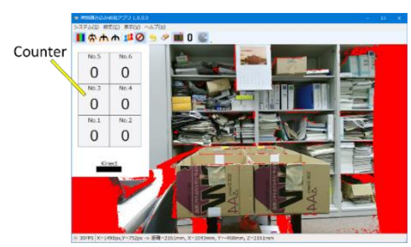

Figure 14. Count application for detecting assortment operation

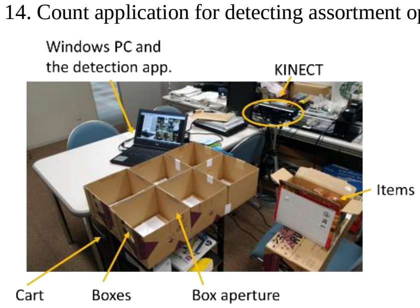

Figure 15. Experimental equipment for delivery

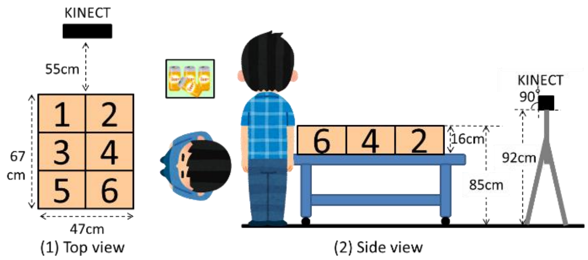

Figure 16. Layout between a worker, MS-KINECT, and assortment of boxes

TABLE VII. EXPERIMENTAL RESULTS OF RECOGNITION IN ASSORTMENT OPERATION

Participant	Number of detectiing correct operation for each box					
	1	2	3	4	5	6
A	5	5	5	5	5	5
B	5	5	5	5	5	5
C	5	5	5	5	5	5
D	5	5	5	5	5	5
E	5	5	5	5	5	5
F	5	5	5	5	5	5
G	5	5	5	5	5	5
H	5	4	5	5	5	5
I	5	5	5	5	5	5
J	5	5	5	5	5	5
K	5	5	5	5	5	5
L	5	5	5	5	5	5
Det. rate	100%	98%	100%	100%	100%	100%

VI. LIGHTING A TAG AND NUMBER

Since our partner AIOI System Co. Ltd. has developed the new lighting method into which the projection mapping technology is used to indicate a picking cell, our experimental system uses this projection mapping technology. A very short focal projector mounted near the MS-KINECT lights a tag attached to a cell and projects a digit on it to indicate the cell and the number of items to be removed as shown in Fig. 17.

A computer knows which tag of a cell is lit, so there is no need for it to detect which tag is lit with the MS-KINECT. In this case, the MS-KINECT is connected to a computer. However, we plan to develop a picking robot that picks items up and puts them into an indicated box in the near future. Since the robot must detect which tag is lit and read a digit on it, we developed a technology that realizes the above functions with the MS- KINECT. In this system, the font used for digits is a seven-segment font as shown on the right side of Fig. 17. Our system recognizes which kind of number is presented by detecting which segments are white.

We noticed that the color through the video camera of the MS-KINECT was very different from the color we recognized and that the color through the video camera of the MS-KINECT changed in accordance with the color and luster of a tag. Example colors on a sheet of white paper, black paper, and gray sandpaper are shown in Fig. 18. The differences between the colors as displayed on a smartphone and those as displayed on the MS-KINECT are shown in Fig. 19. The colors displayed on a smartphone are almost equal to those seen with the naked eye. We selected red, green, and blue as the colors projected on a tag. The color characteristics of the MS- KINECT are very different from those of a smartphone. As a result, gray sandpaper is the best material for representing original colors. Our system can read every number perfectly on a red, green, or blue background. When implementing systems, these three colors are usable.

We think the reason that colors through the MS-KINECT is different from those through the naked eye or displayed on a smartphone is that the white balance of MS-KINECT would be unbalanced. Unfortunately, since the driver of MS-KINECT does not provide adjustment function for the white balance, we recommend persons who use the MS-KINECT as same as our usage to change a material of tag.

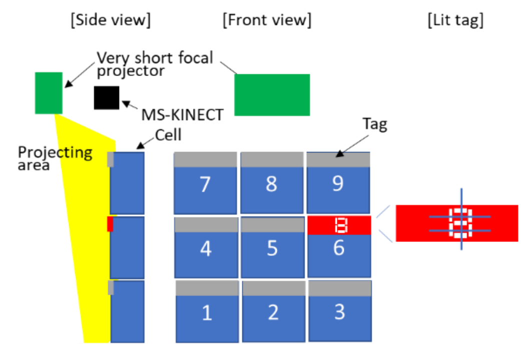

Figure 17. Layout of projector and cells

Figure 18. Colors on tags as displayed on video camera of MS-KINECT

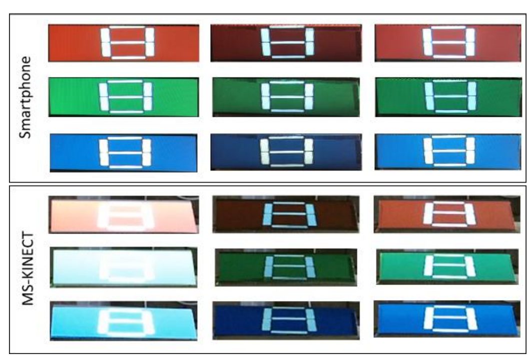

Figure 19. Colors as displayed on smartphone and MS-KINECT

VII. CONCLUSION

One of realistic solutions for establishing high productivity and quality for the picking process in light of a shortage of high-skilled workers is to introduce a picking assistance system that detects incorrect operations by workers. We introduced a picking and assortment assistance system in which the depth camera detects whether a hand is inserted into the correct cell of a shelf to pick items and whether a hand is inserted into the correct box on a cart to put items in.

The misdetection rate for wrong operation is low in these systems, and it is possible to be exceedingly close to zero in a near future. We determine that the proposed picking error detection function would be useful for business.

In this paper, MS-KINECT is used as the depth camera. However, Microsoft stopped to provide MS-KINECT. The other hand, Intel provides the RealSense [15] as the depth camera. That would be usable as the picking and assortment operation assistance systems.

A marketing manager in our partner company, Aioi Systems Co. Ltd which develops and provides picking assistance systems, evaluated the proposed systems. However, he requests systems that do not require a PC for each shelf or delivery cart and a system that can detect incorrect operations by several workers simultaneously. We plan to develop a system in which a stereo-camera is used instead of a MS-KINECT device; results in this paper are applied.

ACKNOWLEDGEMNT

Thanks to Kazuhiro Yoshida for help in performing this research. This project is partially supported with the competitive research funding by Iwate Prefectural University.

REFERENCES

[1] Y. Takehara, Y. Murata, T. Yoshikawa, “Picking Assistance System with MS-KINECT and Projection Mapping,” IARIA, ACHI 2018, pp. 219-223, March 2018.

[2] Industrie 4.0; http://www.bmwi.de/Redaktion/EN/Dossier/industrie-40.html, [retrieved: Augugst, 2018].

[3] What is a smart factory?; http://www.bmwi.de/Redaktion/EN/FAQ/Industrie-4-0/faq-industrie-4-0-03.html, [retrieved: November, 2018].

[4] AutoStore; https://www.autostoresystem.com/, [retrieved: November, 2018]

[5] Introducing the Ocado Smart Platform automated fulfilment solution (promo video); https://www.youtube.com/watch?v=iogFXDWqDak, [retrieved: November, 2018].

[6] Amazon Warehouse Robots: Mind Blowing Video, https://www.youtube.com/watch?v=cLVCGEmkJs0, [retrieved: November, 2018].

[7] AIOI SYSTEMS Co., LTD Digital Picking System Products Introduction; https://www.hello-aioi.com/en/product/product-list/, [retrieved: November, 2018].

[8] Meet Kinect for Windows; https://developer.microsoft.com/en-us/windows/kinect, [retrieved: November, 2018].

[9] Y. Uda, K. Yoshida, and Y. Murata, “Operation Error Detection System in Picking Processes by Camera Depth Sensors,” IPSJ, Transaction of Consumer Device and Service, Vol. 6, No. 1, pp. 1-12, 2016, [written in Japanese].

[10] B. Schwerdtfeger, and G. Klinker, “Supporting order picking with augmented reality,” Proceedings of 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, pp. 91–94, 2008.

[11] H. Iben, H. Baumann, C. Ruthenbeck, and T. Klug, “Visual based picking supported by context awareness: Comparing picking performance using paper-based lists versus list presented on a head mounted display with contextual support,” Proceedings of 11th International Conference on Multimodal Interfaces, pp. 281–288, 2009.

[12] H. Baumann, T. Starner, H. Iben, and P. Zschaler, "Evaluation of graphical user-interfaces for order picking using head-mounted displays," Proceedings of 13th International Conference on Multimodal Interfaces, pp. 377–384, 2011.

[13] A. Guo et al., “A comparison of order picking assisted by head-up display (HUD), cart-mounted display (CMD), light, and paper pick list, “Proceedings of 2014 ACM International Symposium on Wearable Computers, pp. 71–78, 2014.

[14] Open-CV; https://opencv.org/, [retrieved: November, 2018].

[15] Intel RealSense; https://realsense.intel.com, [retrieved: November, 2018].