EVALUATION OF FINGER DIRECTION RECOGNITION METHOD FOR BEHAVIOR CONTROL OF ROBOT

T. Ikai[1], M. Ohka[1], S. Kamiya[1], H. Yussof[2] and S. C. Abdullah[2]

[1]Graduate School of Information Science

Nagoya University Furo-cho Chikusa-ku, Nagoya 464-8601, Japan

[2]Center for Humanoid Robots and Bio-Sensing (HuRoBs), Faculty of Mechanical Engineering,

Universiti Teknologi MARA, Malaysia

Emails: ikai.takuya@b.mbox.nagoya-u.ac.jp; ohka@is.nagoya-u.ac.jp; hanafiah1034@salam.uitm.edu.my

Abstract-When a human gives an order to a robot, the robot must often use its vision to ascertain the human's indication. In our previous paper, to develop a system where robots precisely receive and obey human orders in daily work spaces, we proposed an experimental system for finger direction recognition (FDR) in 3D space using stereo matching by mounting two cameras on the robot. In this paper, we evaluate this system for FDR in 3D space by performing a series of evaluation experiments using a turntable capable of fixing a hand in a specific finger direction. We estimated various finger directions and distances for two major pointing hand forms (Forms 1 and 2) and evaluated the precision of θ and Φ. We conclude that the θ and Φ estimations are valid because the estimation error is almost within 10° when the distance between the camera and object is less than 110 and 80 cm for θ

and Φ for Forms 1 and 2, respectively. Finally, we applied our FDR system to the communication between a robot and a person by visual and tactile sensations. In the application test, the robot recognized the object at which the person pointed and put it in the person's palm.

Index terms*: Robot vision; Human indication; Finger direction recognition; Two hand forms; Handing over

I.   INTRODUCTION

The demand for robots continues to spread in industry and in daily life. We expect the disabled to be able to command robots by such easy actions as pointing and eye contact. In this paper, we introduce a method for conveying human intentions to robots by vision and expect that it will be applied not only to care for the disabled but also to general daily situations.

In order to perceive human behavior, there are various methods such as detection of body acceleration [1] and analysis of video camera [2]. Especially, we make a robot receive a human order, utilizing their vision system is an effective method [3] as many studies of gesture recognition already exist [4]-[11]. Additionally, utilizing vision offers advantages for acquiring an object's color, location, configuration, etc. Using the many advantages of vision for object recognition and receiving human orders, a robot can simultaneously accept the human indications and the indicated objects. This means a strong possibility exists that robots can successfully perform advanced and complex tasks. We adopt a pointing gesture to convey orders to robots.

Pointing gestures, which are often used in daily situations, focus on objects in 3D space from two kinds of information: finger direction and fingertip. Currently, many methods recognize finger direction. One typical method is recognition from the fingertip's position and the finger's base [12]-[14]. However, even if we follow the above studies, recognizing finger direction is difficult when the finger's base is hidden. Other methods utilize the geometric conditions of human hands [15][16]. However, for applications to geometric conditions, since the entire hand must be clearly found in a picture plane, recognizing free hand positions is difficult. Nickel & Stiefelhagen presented an approach for recognizing pointing gestures [17]. For recognition, they visually tracked the head, the hands, and the head orientation but not the actual finger directions.

Recently, Kinect was proposed to accompany new pose detection algorithms [18]. In the research realm of gesture recognition, studies with it are increasing. However, adding Kinect to a robot

requires additional payload, additional free space, and more cost. From the viewpoint of general purposes, we adopt stereo vision as a finger detection sensor.

In our previous study [19], we proposed a new method for finger direction recognition (FDR) in 3D space. In this method, the robot recognized human finger directions without the position information of the base coordinates of the finger or the geometric conditions of the hands using two USB cameras. We defined two finger directions, which were estimated using the position information of the fingertip and the finger side lines to perform FDR in 3D space. One is finger direction θ on the photographic surface; the other is finger direction Φ of the depth. Although we evaluated the FDR method through feasibility experiments, more precise experiments are required.

In this study, to evaluate FDR in 3D space, we performed a series of experiments using a turntable capable of fixing a hand at a specific finger direction. To evaluate finger directions θ and Φ under different hand views, we estimated various finger directions and distances at two hand pointing forms: Forms 1 and 2 (Fig. 1). Finally, in the application test to evaluate FDR in the real world, the robot recognized an object at which the person pointed and put it in the person's palm.
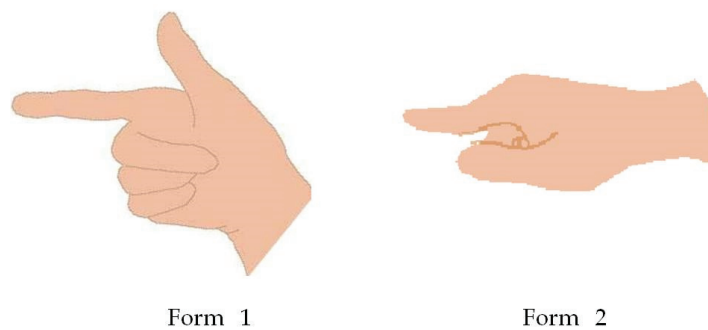

Form 1   Form 2

Figure 1. Pointing forms for evaluation experiment

II.   EXPERIMENTAL ROBOT SYSTEM

In this study, we used a robot head equipped with two USB cameras (Fig. 2) and a personal computer on which OpenCV was installed to process the image data [20]. OpenCV was adopted for hand tracking and utilizes skin detection and stereo calibration. The image data obtained from the two cameras are used for stereo matching, which measures the distance from the camera to

the objects. For stereo matching, the two cameras are set in parallel on the robot's head; the interval between them can be changed freely.

In application tests, we installed this robot head on a two-hand-arm robot [21]. Since three-axis tactile sensors are equipped on its fingertips, the robot can feel the three-dimensional force applied to the grasped objects. In the application test, the robot hands an object at which the operator pointed to the operator.
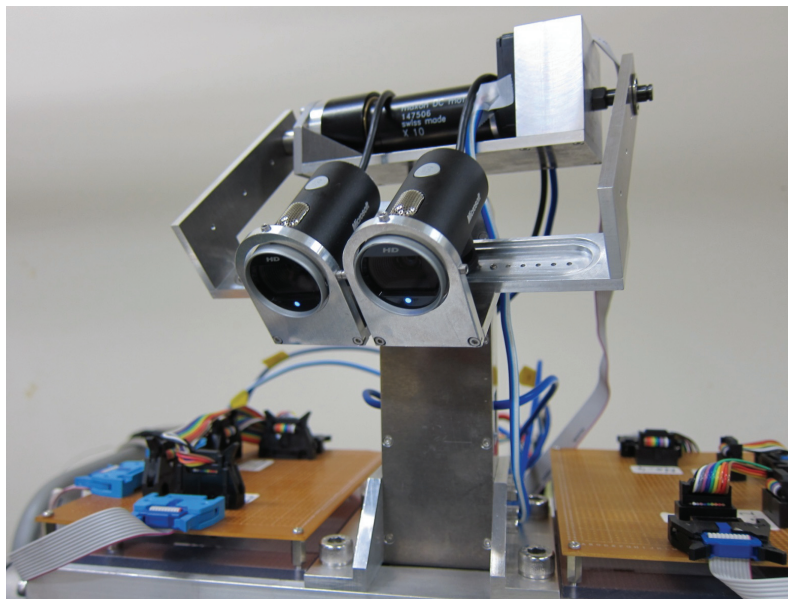

Figure 2. Robot head equipped with two cameras

III.   EXPERIMENTAL PROCEDURE

Although we previously published the FDR algorithms [19], we show them again in the Appendix and include additional explanations, equations, and figures for greater understanding.

Using our algorithms, we made an FDR system for 3D space that can work about three frames per second. To evaluate its accuracy, we made a turntable to fix the hand that has two axes that move freely in rotation directions θ and Φ. Since it does not have any colors that resemble red, it does not interfere with finding the hand region.

In this experiment, we define the polar coordinates of the finger direction in Fig. 3.
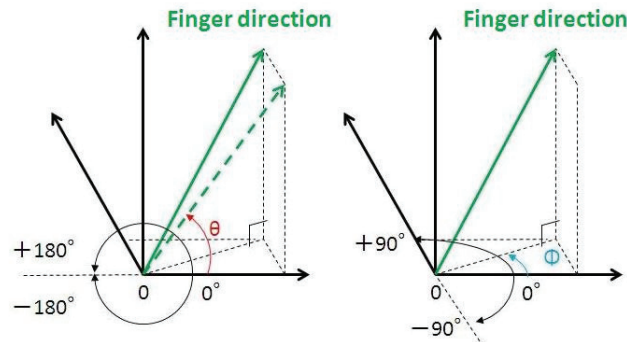
Figure 3. Angle definitions of θ and Φ

We measured the precision of estimating θ and Φ. We put a hand away from the camera and performed a pointing gesture using the first finger (Fig. 4); we used a blue background. We performed experiments for two major pointing forms (Fig. 1). In Form 1, we showed the back or palm of the hand. In Form 2, we hid all of the fingers except the index finger.

In the experiments, we used the configuration of the cameras and the pointing finger shown in Fig. 4. We estimated θ and Φ for ten seconds and fixed the finger angle during the estimation. We performed this experiment for θ = 112.5, 135, 157.5 and 180° and Φ = -45, 0, and 45°. We performed a similar experiment on different distances from the camera: 50, 80 and 110 [cm]. After the estimation, we compared the demonstrated angles with the estimated angles to evaluate this system.

As another evaluation standard, we also defined the following detection rate (*DR*):

$$DR[\%] = 100 \times \frac{\text{Frames capable of estimating the direction}}{\text{Processed frames}} . \tag{1}$$

We performed the estimation for consecutive still frame images. A high *DR* means that our system estimated the finger direction many times within a specific time. Even if the estimation values are accurate for both θ and Φ, since the number of frames capable of estimating a finger direction becomes more important for transmitting human indications to the robots, the estimation accuracy is not always useful in real time.
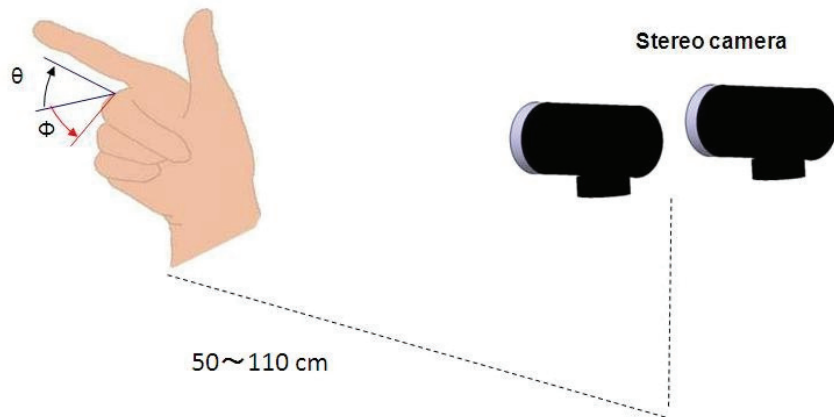
Figure 4. Evaluation experiment

IV.   EXPERIMENTAL RESULTS AND DISCUSSION

a. Experimental results

Tables 1A and 2A in the Appendix show the experimental results with FDR. Figs. 5-9 are structured graphs that show the accuracy and the standard deviation at each distance. On the whole, θ is superior to Φ in the accuracy of the estimation values, and Φ is superior to θ in the detection rate (DR).

If we check the DR values in Tables 1A and 2A, we notice high DR values in the Φ estimation and very low values in the θ estimation. This inclination is not different between Forms 1 and 2. In these tables, values less than 50% are underlined. Since these underlined values are marked in the θ estimation, the Φ estimation is superior to the θ estimation. The FDR method can be used even in low DR because the computer processing is done quickly enough. However, we must improve FDR to get a higher DR.

Figure 5 shows the estimation of θ of Form 1. Figs. 5(a), (b), and (c) are the 50-, 80-, and 110-cm distance cases, respectively. Each graph includes three Φ conditions: -45, 0, and 45°. Although these graphs include such irregular data as the points of $\Phi = 45°$ and $\theta = 110°$ in Fig. 5(b), almost all of the data align to 45° inclination. This means that the indicated angle θ is estimated by the FDR method. Especially in the case of $\Phi = 0°$, indicated angle θ is completely estimated because the data precisely align to a 45°, and the standard deviation (SD) shown by the vertical line is almost zero.
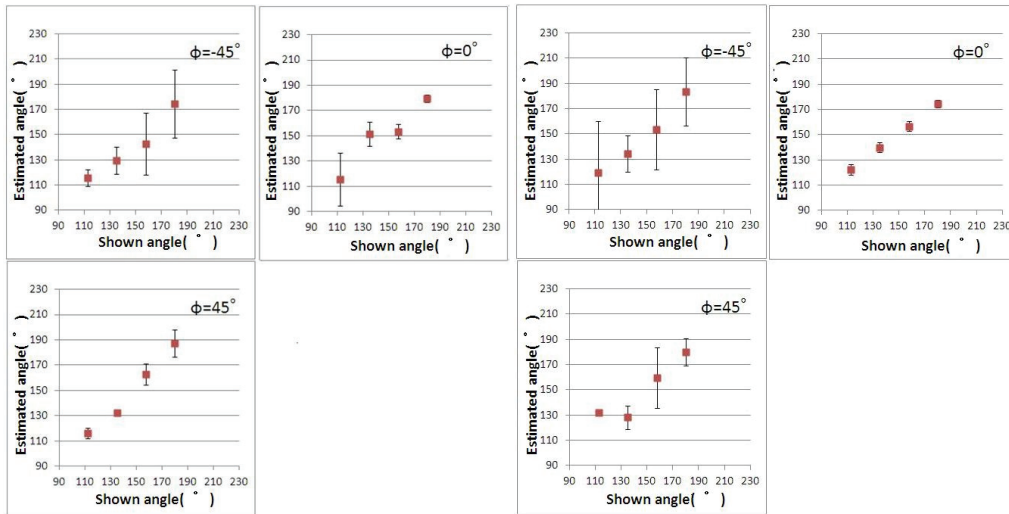
(a) 50 cm   (b) 80 cm
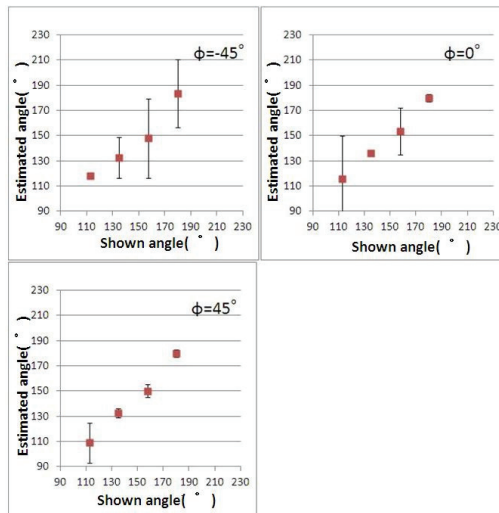
(c) 110 cm

Figure 5. Accuracy and standard deviation of θ at each Φ at each distance (Form 1)

Next, Fig. 6 shows the estimation of Φ of Form 1. Figs. 6(a), (b), and (c) are the 50-, 80-, and 110-cm distance cases, respectively. Each graph includes four θ conditions: 112.5, 135, 157.5, and 180°. Although these graphs also include some irregular data, almost all of the data align to a 45° inclination.
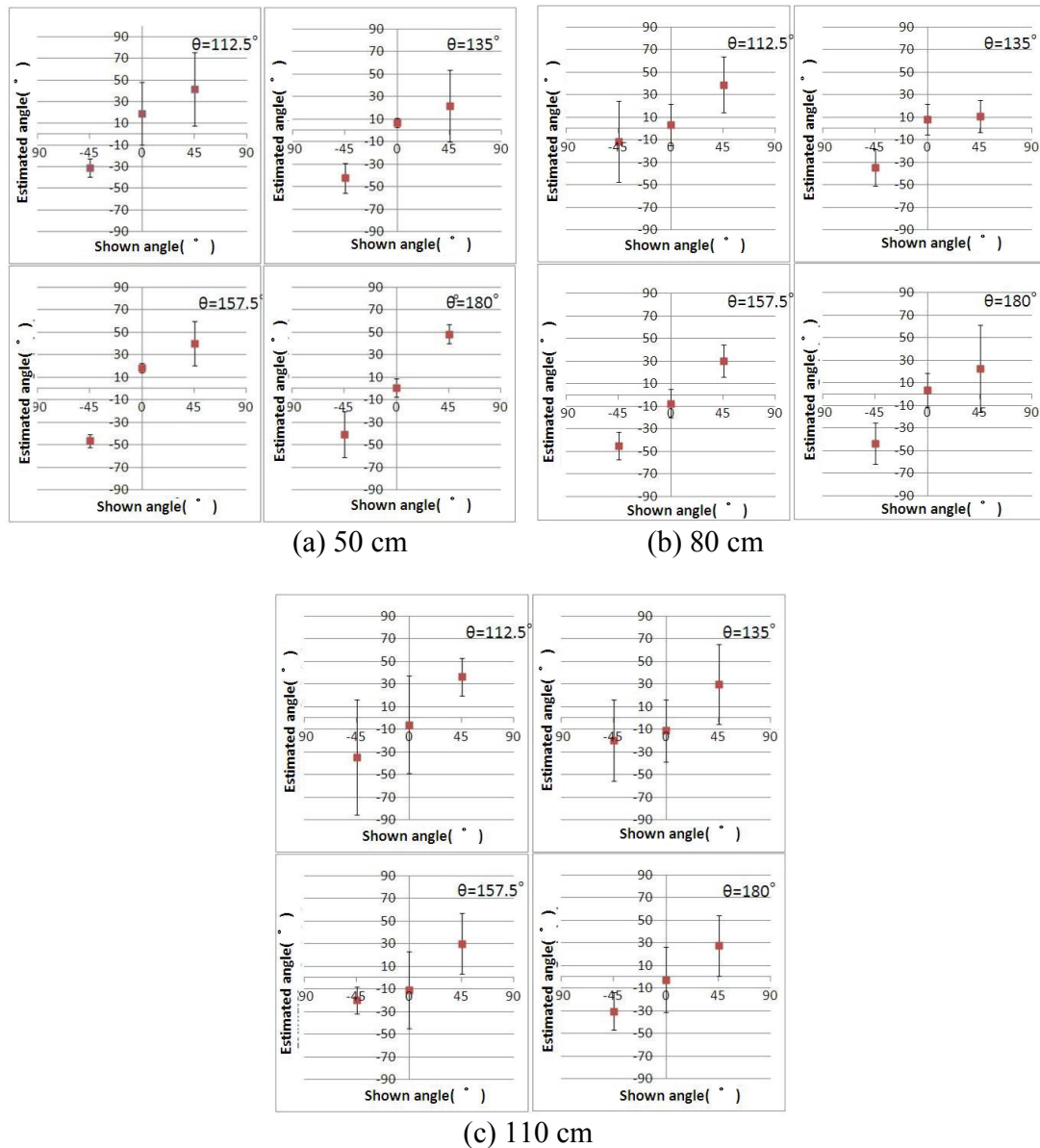
(a) 50 cm   (b) 80 cm

(c) 110 cm

Figure 6. Accuracy and standard deviation of Φ at each θ at each distance (Form 1)

As mentioned for Figs. 5 and 6, the estimations of θ and Φ are well done for Form 1. The precision of Φ is inferior to θ because the SD of Φ is markedly larger than that of θ. Since the distance calculation is included for the Φ estimation in FDR, a one-pixel difference on the stereoscopic images induces a few centimeters of error from 50 to 110 cm.

(a) 50 cm   (b) 80 cm
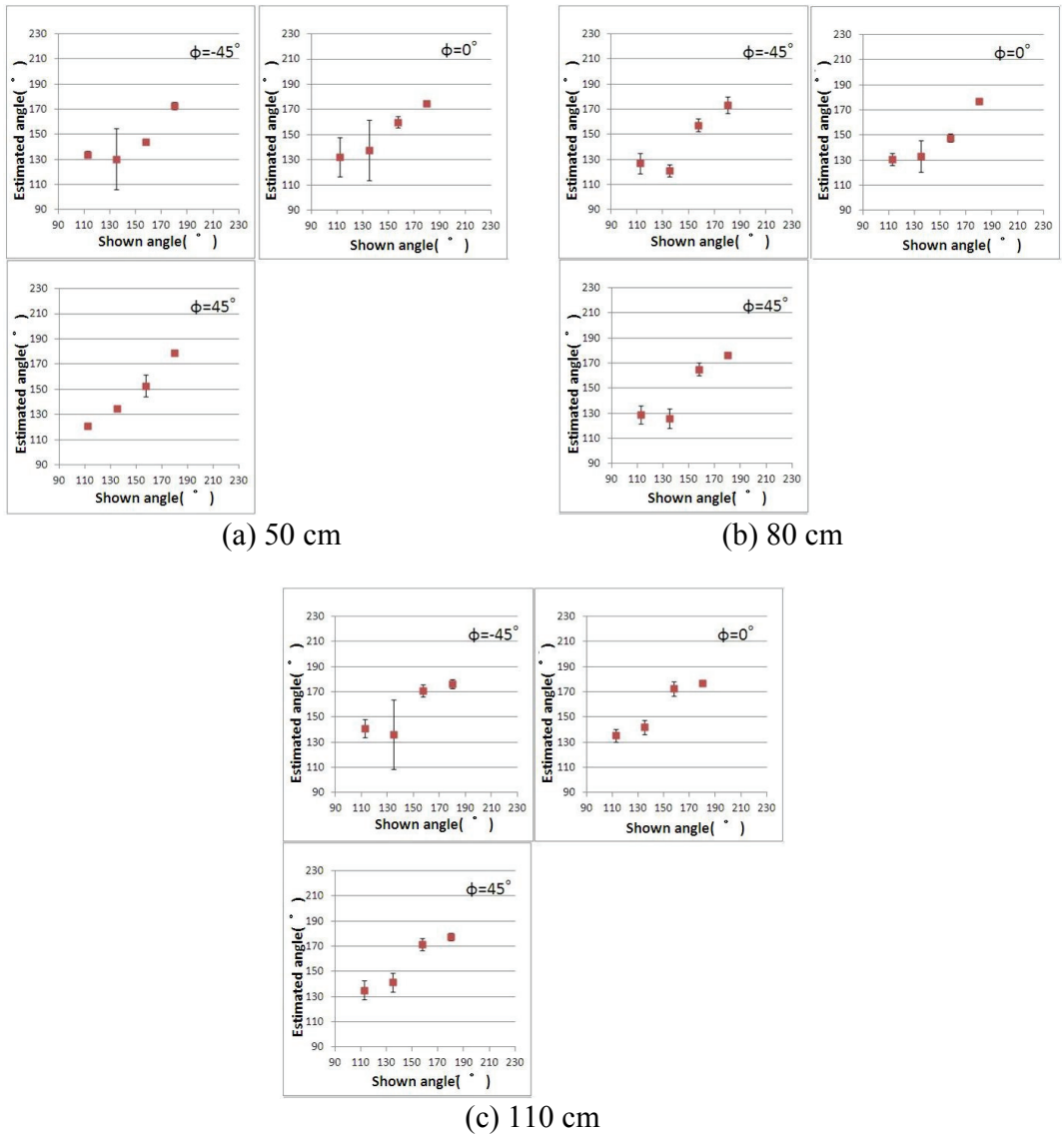
(c) 110 cm

Figure **7.** Accuracy and standard deviation of θ at each Φ at each distance (Form 2)

On the other hand, Fig. 7 shows the estimation of θ of Form 2. Figs. 5(a), (b), and (c) are the 50-, 80-, and 110-cm distance cases, respectively. While almost all of the data align to a 45° inclination, as in Fig. 5, the data in Fig. 7 do not completely align to a 45° line. Since the SD values of Fig. 7 are not bigger than those of Fig. 5, even Form 2 can be applied to the pointing.
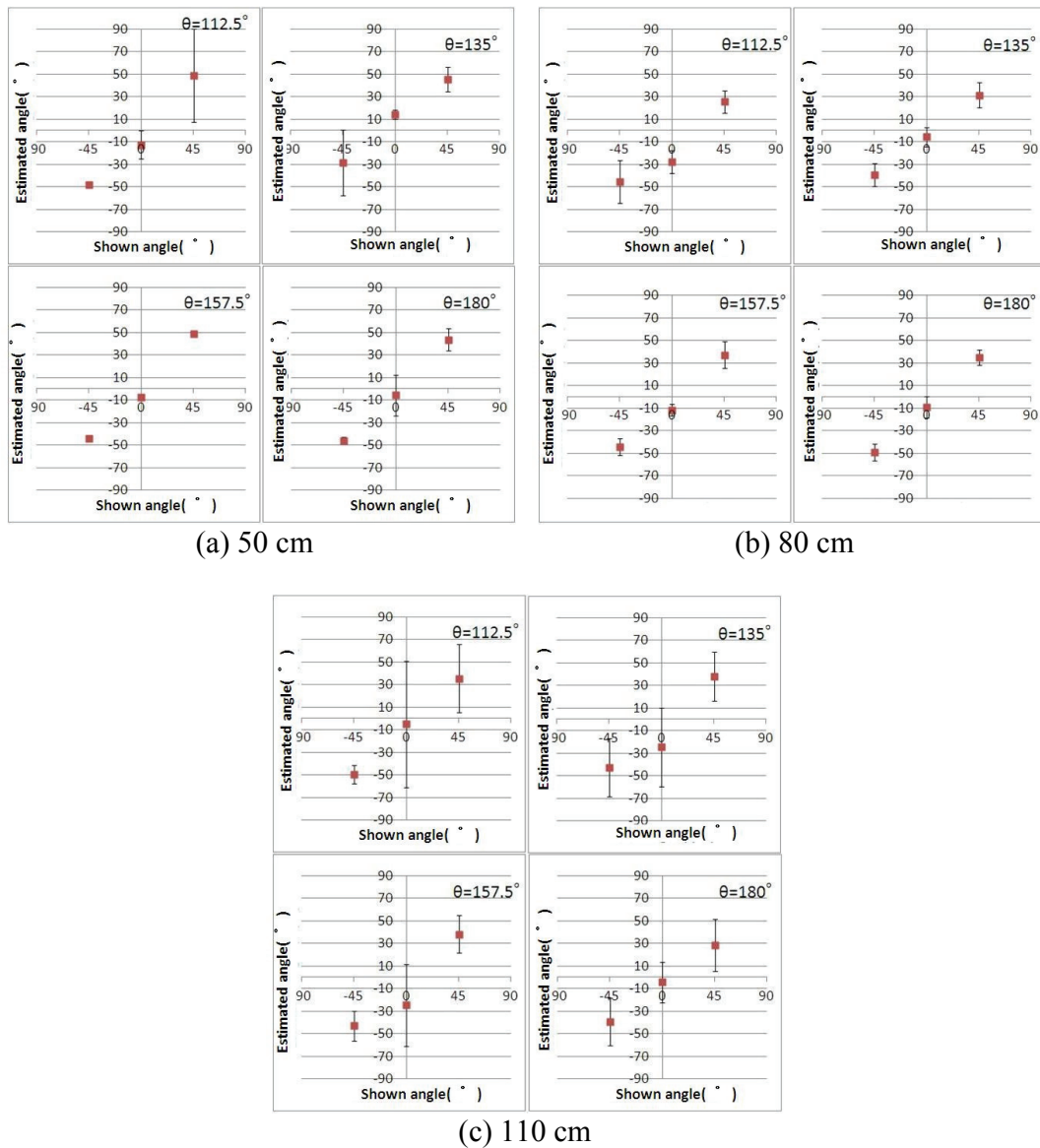
(a) 50 cm

(b) 80 cm

(c) 110 cm

Figure 8. Accuracy and standard deviation of Φ at each θ at each distance (Form 2)

Finally, Fig. 8 shows the estimation of Φ for Form 2. The data in Fig. 8 almost align to a 45° line. Since the SD values of Fig. 8 are smaller than those of Fig. 6, the estimation of Φ for Form 2 outperforms that of Form 1.

b. Discussion about estimations of θ and Φ

In the estimation of θ, Form 1 is superior to Form 2. In the estimation of θ, we utilize two straight lines of the finger side in the image. In Form 1, since these lines are approximately parallel, FDR can accurately decide the finger direction. However, in the case of Form 2, since the finger side

corresponds to the finger's pad and back lines, which are uneven, estimated θ differs little from the true value.

In contrast, in the estimation of Φ, Form 2 is superior to Form 1 because of the accuracy of the estimation values and the smallness of the SD, probably because of the small disparity between the left and right cameras. In Form 2, the middle, ring, and little fingers are hidden behind the index finger and the thumb. On the other hand, in Form 1, these three fingers are more visible, especially at $\Phi = 45°$ and $\Phi = -45°$. Their large volume dominates the hand images and makes different hand shapes between the left and right images. Since their volume causes the hand region's centroid to deviate from the correct center, image processing cannot correctly obtain the disparity that corresponds to the distance. Therefore, the stereo matching accuracy deteriorates and the Form 1 results worsen.

The estimation of θ marks the best value at $\Phi = 0°$ of either form. In this angle, the photographic surface faces the cameras, and we can easily find the finger side lines. Since they are near a finger direction vector, θ marks the best value.

On the other hand, the estimation of Φ marks the best value at $\theta = 180°$ in either form. In other θ angles, some errors seem to be caused by the gap between the finger side lines and the correct finger direction vector. When we point obliquely upward, three fingers (not the index finger and thumb) are more visible. As discussed previously, they worsen the stereo matching accuracy. Although Form 2 marks a good Φ value, the case in which we point obliquely upward is slightly bad for the above reason.

For both forms, in the estimation of θ, we can't determine whether the estimation accuracy is decayed progressively by distance. However, in the estimation of Φ, estimation accuracy does decay progressively with distance. As shown in Fig. 9A of the Appendix, we utilize L (the horizontal distance between the fingertip and the gravity center of the hand region) in the estimation of Φ. Since L [m] is expressed as L' [pixel] in the camera image, the longer the distance to the camera becomes, the longer is the distance expressed by one pixel. Thus, the fingertip position error between the left and right images becomes larger, too. This is one reason that the estimation accuracy is decayed progressively with distance.

c. Application to a two-hand-arm robot

In another project, we developed a two-hand-arm robot [21] that grasps the objects at which an operator points and gives them to the operator. We apply this method to our robot to evaluate the communication method between it and humans in an experimental situation.

Figure 9 shows photographs from our experiment. First, an operator points at a white die to indicate that the robot should pick it up. In Photo 1, the robot recognizes the operator's finger direction using FDR. In Photo 2, it estimates the die position by calculating the union of the finger direction vector and the die image. In Photo 3, the die's destination is determined as the centroid of the largest skin color area. Photos 4 to 7 show the die being picked up and placed on the hand.
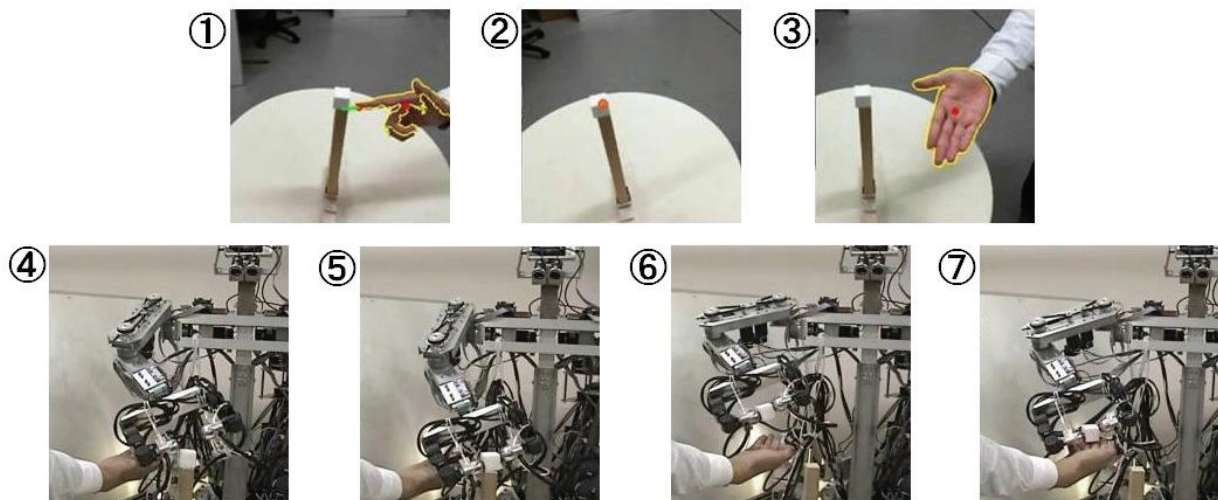

Figure 9. FDR evaluation test using two-hand-arm robot

V. CONCLUSION

We introduced an experimental system for FDR in 3D space using stereo matching with two cameras and mounted it on a robot. This method can be applied to cases when the bases of fingers are hidden. We estimated two finger directions: θ on the photographic surface and Φ of the depth. They are projections of the finger direction in the 3D space to two planes. We applied this method to two major forms (Forms 1 and 2) for pointing. Form 1, which shows the hand's back or palm, is superior for estimating θ. Form 2, which hides all of the fingers (except the index

finger), is superior for estimating Φ. Each of these forms has its merits and demerits, but the accuracy of θ of Form 2 is better than the accuracy of Φ of Form 1. Therefore, we adopt the latter as object pointing.

In this method, we defined a pointing finger as the farthest finger from the hand region's center of gravity. Therefore, this method can be applied not only to the first finger but to the other fingers as well. Although several problems remain to be solved in our proposed method, we showed in this paper that it can work in a controlled environment. If we adjust the program's parameters, it can estimate the finger direction in specific environments. Our future work will address the aforementioned problems to relax the limitations.

In other future work, we will explore an estimator that is robust against light. In this paper, we simply chose red as a skin color. Since several superior skin color detection methods have been introduced in the field of face detection [22], we will experiment with them. Since speech preceding gestures that support gesture extraction is known [23], we would also like to add such recognition. By utilizing the abovementioned technics, robots will be able to recognize a human's finger in various situations. Finally, we will endeavor to make a robot grasp objects to which a human is pointing.

ACKNOWLEGMENT

APPENDIX

IA. EXTRACTION OF HAND AREA

When we accomplish FDR, we have to determine hand area on the image first. The human hand is usually skin color composed of red and yellow components. By excluding other colors from an image, we can get only the skin-color of hand and face regions. Additionally, masking the face region, which has skin color, we can help obtain the hand region. Already, there are many studies on obtaining the hand area and fingertip point on an image; based on these ideas, we obtain the hand area on the image and fingertip point [24][25].

a. Extraction of Hand Contour

First, we extract only red areas from an image and masked face regions using OpenCV (Figure 1A). Second, we make those areas' binary image and regard the biggest area from them as the hand area and obtain its contour.
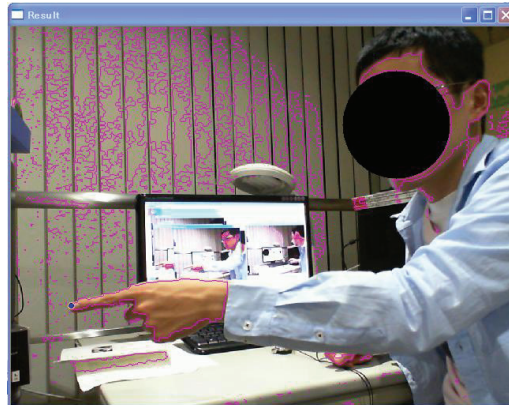

Figure 1A. Example of masking face and extracting red area

b. Identification of Fingertip

A fingertip is specified as a pointed place in the hand area and has a big curvature. Here, we regard contours as an aggregate of contiguous pixels and pursue a pixel trajectory.

Along the contour, we check how sharp convex in contour each pixel is and regard the convex whose sharpness exceeds a threshold as the fingertip (Figure 2A). To check how sharp a target convex is, we utilize its anterior-posterior points and vectors formed by these three points. Now we define a target point $P_{target}(x_{target}, y_{target})$ and property near points, $P_a(x_a, y_a)$ and $P_b(x_b, y_b)$. Then, we assume two vectors $\vec{A} = \overrightarrow{P_{target}P_a}$ $\vec{B} = \overrightarrow{P_{target}P_b}$, and define the angle between $\vec{A}$ and $\vec{B}$ as Θ_α (Figure 3A). In the place such as a fingertip, where the tangential direction of a curve line causes considerably large change, Θ_α becomes small. This Θ_α depends on the values of $P_a(x_a, y_a)$ and $P_b(x_b, y_b)$, so we check Θ_α about some cases that $P_a(x_a, y_a)$ and $P_b(x_b, y_b)$ are distant from 5 pixel to about 20 pixel. And we choose the smallest Θ_α among them and define its value as sharpness of $P_{target}(x_{target}, y_{target})$. In this way, we calculate sharpness value of each pixel, and regard the point whose sharpness exceeds a threshold as fingertip. If those points

are in succession, we regard them as one fingertip and choose only one point which has smallest Θ_α among them.

In above method, we compare smallness of Θ_α. At that time, we utilize an inner product $\vec{A} \cdot \vec{B}$ (where, $\vec{A} = (a_x, a_y)$, $\vec{B} = (b_x, b_y)$) as the magnitude index. For example, when Θ_α becomes smaller, $\vec{A} \cdot \vec{B}$ takes a bigger value. In addition, because a fingertip takes the form of an acute angle, we utilize outer products $\vec{A} \times \vec{B}$ to distinguish between an acute angle and an obtuse angle. If component of $\vec{A} \times \vec{B}$ has positive value, Θ_α is an acute angle. Thus, the conditions to keep the point on the hand contour as a fingertip are

$$\cos\Theta_\alpha = \max \frac{a_x b_x + a_y b_y}{|\vec{A}||\vec{B}|} \qquad (1A)$$

$$a_x b_y - a_y b_x > 0 \qquad (2A)$$

We can find fingertips by using the above two formulas.

c. Identification of Center of Gravity of Hand Region

Finally, we calculate the position of the center of gravity of the hand region. We need this point for estimation of finger direction Φ. It is calculated easily from the binary hand contour. Figure 4A shows the result of above estimation process. After the contour of hand is obtained from the original image, the fingertip of pointing finger and the center of gravity are indentified.
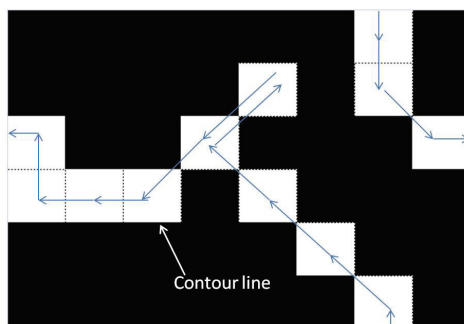

Figure 2A. Purse contour line (at all pixel, we calculate the sharpness in contour.)
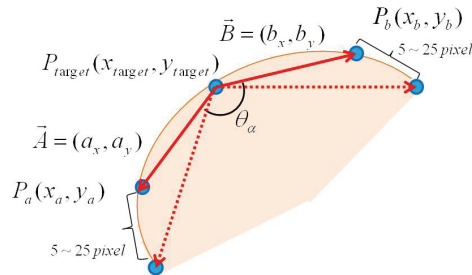
Figure 3A. Θ_α on finger contour (we chose $P_a(x_a, y_a)$ and $P_b(x_b, y_b)$ inside a certain range and check whether the sharpest Θ_α exceeds a threshold or not.)
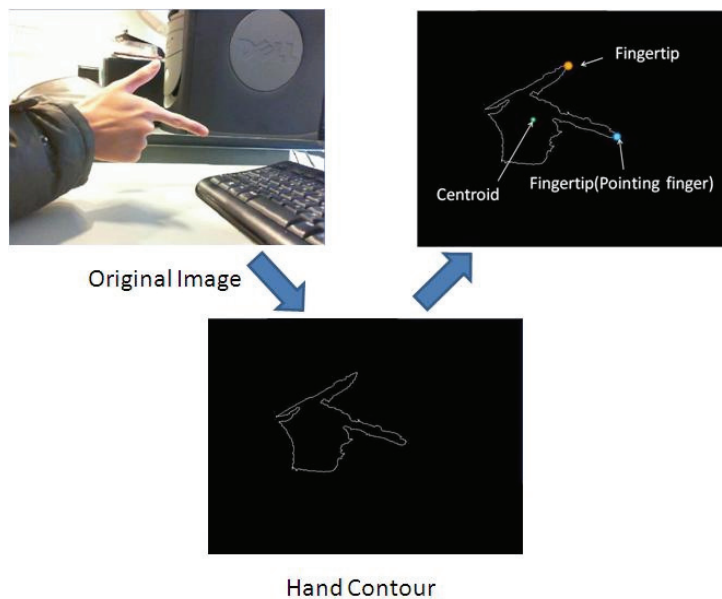
Figure 4A. Estimation process of finger direction

IIA. ESTIMATION OF FINGER DIRECTION IN ELEVATION

First, to obtain the optimum condition of wire distance, we surveyed the effects of various wire Finger direction θ is the projection of finger direction in 3D space to photographic planes. Figure 5A shows the relationship between 3D finger direction and Finger direction θ.

Finger direction θ can be determined as the average angle of two straight lines of the finger side. If we assume unit vectors along both sides of the finger, the finger direction vector in the photographic plane is defined as the sum of them (Figure 6A).

To detect lines of the finger side, we detect straight lines using the Hough transform. Detecting straight lines by the Hough transform is already used to detect finger lines. However, since the

Hough transformation consumes much computational time, it is inadequate for real time estimation. To overcome this problem, detecting straight lines only around the fingertip makes real time estimation possible.

The method to estimate finger direction θ is composed of the following five steps in the procedure. These steps are repeated in every image data processing. Figure 7A shows results of estimation based on the above procedure.

(i)   Retrieve only the hand area from an image and make the retrieved area binary data.

(ii)   Cut an image in a square shape and retrieve only a limited area of the fingertip (the center of this cutting window image coincides with the fingertip).

(iii)   Detect straight lines using Hough transformation and lines of the finger side in the cut image. If there are some straight lines, decide unit vectors on the line.

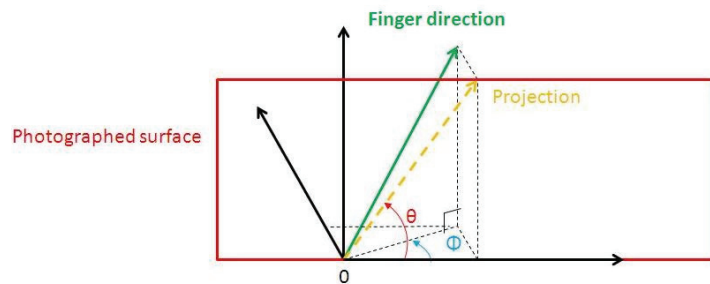(iv)   Sum up all unit vectors and decide finger direction.

Figure 5A. Finger direction on photographed surface
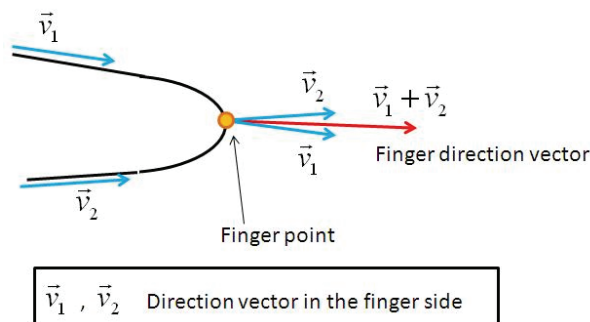
Figure 6A. Mechanism of estimating finger direction vector
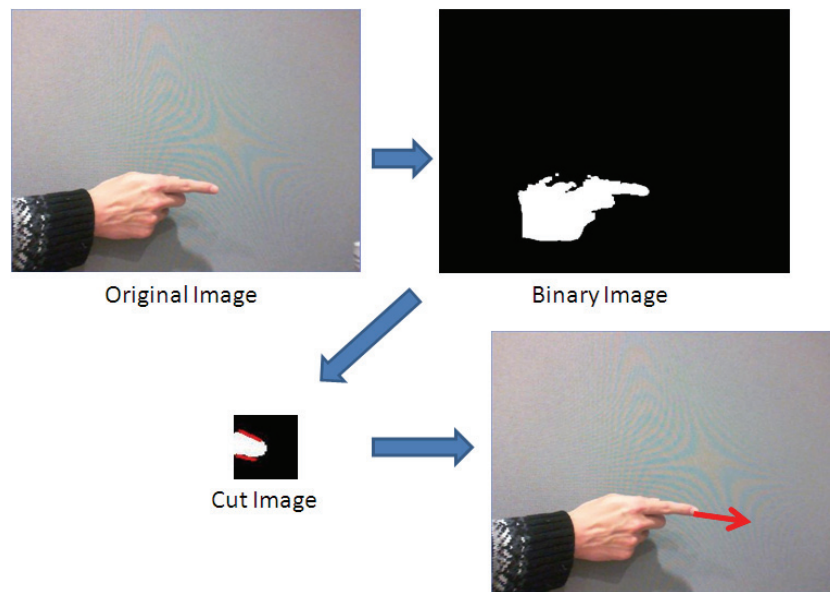
Figure 7A. Finger direction on photographed surface

IIIA. ESTIMATION OF FINGER DIRECTION IN TOP-VIEW

In this section, we describe the estimation of finger direction in top-view Φ using stereo matching. In this estimation, we use both right and left camera images.

a.   Definition of Finger Direction in Top-View

Direction in top-view Φ of depth means the angle between the hand and the photographic surface as shown in Figure 5A. To estimate Φ, we calculate 3D position of two points: the fingertip being used on a pointing action and the centroid gravity of the hand region. Here L is designated to the horizontal distance between the fingertip and the centroid of the hand region. Also, distances from a camera to the fingertip and to the centroid of the hand region are D_G and D_F, respectively. Finger direction Φ can be estimated by means of the following formula.

$$\Phi = \tan^{-1} \frac{D_G - D_F}{L} \tag{3A}$$

Relationship among these parameters is shown in Figure 8A. D_F and D_G are obtained from the stereo matching. Since L is displayed as L'[pixel] on the camera image (Figure 9A), it is calculated from L' if we know the camera parameters.
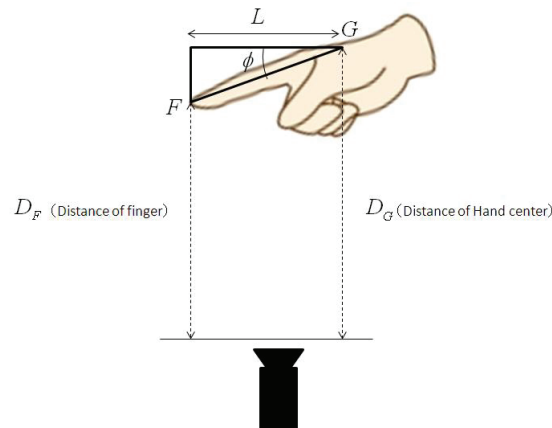
Figure 8A. Estimation of depth pointing angle Φ
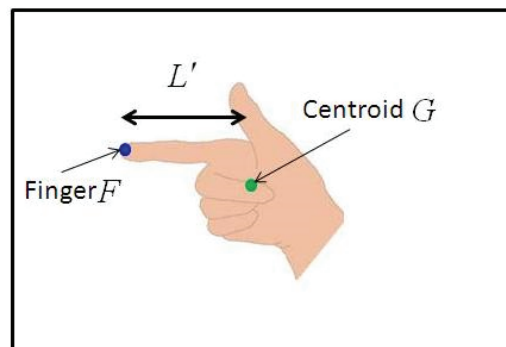
Figure 9A. Definition of L′ on camera image

b.   Stereo Matching

Here, we explain our measuring technique for finger direction in top-view Φ that takes advantage of stereo matching. The stereo camera system possesses a common Z-axis and each camera in the system has each local X-axis. If we compare two images obtained from cameras, we know the nearer object has larger disparity between two images. This disparity is proportional to the distance to the object and is a clue to estimating the distance to the object.

We name the each position of the object on the right and left image x_R and x_L. In addition, we name the distance, which is called base length B and the focal length f [pixel].

The relationship between these parameters and the distance Z to the object are represented as the following formula.

$$|x_L - x_R| : f = B : Z$$

$$Z = \frac{Bf}{|x_L - x_R|} \qquad . \tag{4A}$$

We use this formula to estimate the distance to the object.

As mentioned, the parameter of the camera is essential in the stereo matching. In the calculation, since the focal length f is usually expressed in metric units, we must convert it into pixels. In addition, we must modify errors of the equipment. In this section, we formulate an actual formula from the above theoretical formula.

If we introduce a new parameter k ($k = Bf$) and substitute it into Eq. (4A),

$$Z = \frac{k}{|x_L - x_R|} \tag{5A}$$

Since the unit of B is [m] while the unit of f is [pixel], the unit of k is [m·pixel]. Next, we assume stationary error as e_1, e_2, e_3.

$$Z = \frac{ke_3}{|x_L - x_R| - e_1} - e_2 \tag{6A}$$

We designate K to ke_3. Finally, Eq. (6A) is deformed to

$$Z = \frac{K}{|x_L - x_R| - e_1} - e_2 \tag{7A}$$

This formula is applied to estimate the distance to the object. K, e_1 and e_2 are unknown parameters and we must evaluate these parameters by preliminary experiments as follows.

c. Estimation Procedure

We set parallel cameras on the tripod and face them toward a wall. Cameras are put on specific positions, which are Z[m] away from a wall marked with a dot. We compare the mark position between right and left images, and calculate disparity $|x_L - x_R|$ [pixel]. From the relationship between Z and $|x_L - x_R|$, we can obtain the value of unknown parameters.

Next, we describe a transformation from L'[pixel] to L[m]. A convex lens of a digital camera focuses into an image on a CCD (Charge Coupled Device) image sensor. On CCD, many image pickup devices are evenly distributed and each of the image pickup devices corresponds to one pixel on an image. Accordingly, if we want to know the actual size of an object from a digital image, we should ascertain the relationship between the size of an actual object and the size of an object image projected through a convex lens. In other words, we should get a magnification of the lens. Convex lenses follow a simple equation that determines the location of the images (a), the location of the images (b), the size of the images (L) and the size of the images (L_p):

$$L : a = L_p : b$$

This formula is deformed as

$$L = \frac{a}{b} L_p \tag{8A}$$

The magnification of a lens is expressed by $\frac{a}{b}$. Here, the unit of L and L_p is [m]. By a preliminary experiment, we can know the relationship between L and L_p.

The method for estimating finger direction Φ follows five steps. This procedure is repeated every frame.

(i) Compare the right and left images and calculate the disparity of fingertips.

(ii) As with (i), calculate the disparity of the center of gravity of the hand region.

(iii) By using stereo matching, calculate the distance from cameras to fingertip and the distance to the center of gravity of the hand region.

(iv) Calculate L' (the horizontal distance between a fingertip and the center of gravity of the

hand region on an image). Then, convert L' to an actual distance L.

(v)   Determine finger direction Φ using D_F, D_G and L.

Table 1A.   Result of evaluation experiment (Form 1)
(a) 50 cm   (b) 80 cm

Distance   50cm

$\theta(°)$	DR(%)	Average	SD	$\Phi(°)$	DR(%)	Average	SD
112.5	34.6	115.4	20.8	0	79.5	19.1	28.8
135	21.4	151.3	9.7	0	98.8	6.8	3.9
157.5	41.7	153.4	5.7	0	89.1	18.2	4.3
180	59.8	179.7	3.1	0	98.8	0.5	8.4
112.5	15.6	115.8	6.6	45	64.1	41.3	8.6
135	43.6	129.4	10.7	45	95.7	21.5	32.1
157.5	50.6	142.6	24.5	45	91.6	39.6	19.6
180	56.2	174.4	26.8	45	98.8	48.3	8.3
112.5	16.7	116.3	4.4	−45	78.5	−31.2	33.9
135	50.6	132.2	1.9	−45	95.9	−42.4	13.2
157.5	42.1	162.9	8.5	−45	99.2	−46.4	5.8
180	65.4	187.3	10.7	−45	100.0	−40.7	20.6

Distance   80cm

$\theta(°)$	DR(%)	Average	SD	$\Phi(°)$	DR(%)	Average	SD
112.5	50.8	122.4	4.3	0	76.1	3.2	18.3
135	91.0	139.8	3.9	0	99.7	7.8	13.7
157.5	79.7	156.6	3.9	0	86.9	−7.7	12.6
180	48.0	174.4	3.1	0	50.8	3.7	15.2
112.5	38.1	119.4	40.6	45	85.2	38.6	24.7
135	49.9	134.3	14.5	45	81.8	10.7	14.3
157.5	65.6	153.5	32.0	45	93.8	30.2	14.1
180	39.8	183.6	26.8	45	49.2	22.8	38.5
112.5	13.2	132.2	1.9	−45	78.9	−11.6	36.0
135	67.3	128.0	9.1	−45	70.2	−34.7	16.0
157.5	83.4	159.5	24.2	−45	90.0	−45.3	12.2
180	86.9	179.7	10.7	−45	27.4	−43.9	18.3

(c) 110cm

Distance   110cm

$\theta(°)$	DR(%)	Average	SD	$\Phi(°)$	DR(%)	Average	SD
112.5	42.0	115.5	34.5	0	71.5	−5.9	42.9
135	41.3	136.3	2.2	0	100.0	−11.0	27.4
157.5	100.0	153.6	18.5	0	91.3	5.6	33.9
180	79.8	179.7	3.1	0	93.9	−2.5	29.1
112.5	14.3	117.9	1.8	45	32.7	36.3	16.6
135	82.2	132.6	16.3	45	94.4	30.0	35.2
157.5	36.2	148.0	31.3	45	100.0	34.4	26.8
180	48.5	183.6	26.8	45	68.4	27.6	26.8
112.5	41.3	108.9	15.9	−45	56.3	−34.6	50.7
135	18.9	132.6	3.6	−45	72.4	−19.8	35.7
157.5	20.0	150.2	5.1	−45	77.6	−42.0	11.8
180	42.5	179.7	3.1	−45	82.5	−30.3	16.8

Table 2A.   Result of evaluation experiment (Form 2)

(a) 50 cm   (b)80 cm

Distance   50cm

θ(°)	DR(%)	Average	SD	Φ(°)	DR(%)	Average	SD
112.5	38.1	132.5	15.5	0	68.0	−12.7	12.4
135	41.0	137.7	24.1	0	72.9	14.3	4.2
157.5	59.2	160.1	4.7	0	95.1	−7.3	2.6
180	64.5	174.7	2.4	0	88.7	−5.6	18.0
112.5	50.9	134.0	2.8	45	98.1	48.7	2.9
135	23.3	130.4	24.1	45	66.7	45.6	10.8
157.5	21.9	144.1	1.1	45	65.5	48.8	2.9
180	19.6	172.6	3.1	45	100.0	43.6	9.8
112.5	29.1	121.1	1.8	−45	59.2	−48.2	41.2
135	66.2	134.9	1.9	−45	86.7	−28.8	29.2
157.5	35.7	152.7	8.6	−45	45.6	−44.1	2.6
180	63.6	178.9	1.7	−45	82.7	−46.0	3.1

Distance   80cm

θ(°)	DR(%)	Average	SD	Φ(°)	DR(%)	Average	SD
112.5	55.1	130.7	4.8	0	81.1	−27.9	10.0
135	55.5	133.0	12.6	0	81.6	−5.7	8.5
157.5	54.1	147.8	3.4	0	80.0	−11.6	5.3
180	41.7	176.8	2.1	0	91.7	−9.1	9.5
112.5	58.1	126.9	8.0	45	82.8	25.4	9.9
135	44.5	121.1	4.7	45	80.5	31.1	11.2
157.5	57.8	157.3	5.1	45	87.9	37.1	11.8
180	26.2	173.3	6.8	45	81.6	35.0	6.7
112.5	62.0	129.1	7.2	−45	83.7	−45.7	19.0
135	66.0	125.8	7.9	−45	88.0	−39.5	10.1
157.5	73.6	165.0	5.2	−45	93.8	−44.5	7.5
180	60.4	176.5	1.0	−45	92.2	−48.9	7.4

(c) 110 cm

Distance   110cm

θ(°)	DR(%)	Average	SD	Φ(°)	DR(%)	Average	SD
112.5	42.1	135.3	5.1	0	91.2	−5.1	55.8
135	61.0	142.1	5.7	0	88.0	−24.8	35.1
157.5	53.7	172.6	5.7	0	72.1	9.9	36.4
180	48.4	176.9	1.6	0	79.0	−4.4	17.9
112.5	21.0	140.8	7.2	45	62.5	35.4	30.3
135	54.1	136.0	27.7	45	44.7	38.1	21.6
157.5	41.9	171.0	4.9	45	85.0	36.0	16.6
180	12.7	176.4	3.6	45	44.1	28.5	22.8
112.5	27.0	135.1	7.7	−45	84.4	−49.7	8.3
135	38.0	141.4	7.5	−45	89.7	−42.9	25.8
157.5	49.2	171.5	4.6	−45	93.3	−51.4	13.3
180	59.3	177.6	2.9	−45	80.2	−39.4	21.0

REFERENCES

[1] Morioka K, Hashikawa F, and Takigawa T, (2013), "Human Identification Based on Walking Detection with AccelerationvSensor and Networked Laser Range Sensors in Intelligent Space," International Journal on Smart Sensing and Intelligent Systems, Vol. 6, No. 5, 2040-2054.

[2] Ahmed HS, Faouzi BM, and Caelen J, (2013), "Detection and Classification of the Behavior of People in an Intelligent Building by Camera," International Journal on Smart Sensing and Intelligent Systems, Vol. 6, No. 4, 1318-1342.

[3] Jain R, Kasturi R, and Schunck BG, (1995), "Machine Vision," McGraw-Hill, Inc., ISBN 0-07-032018-7.

[4] Kurata T, Kato T, Kourogi M, Keechul J, and Endo K, (2002), "A Functionally-distributed Hand Tracking Method for Wearable Visual Interfaces and its Applications," In IAPR Workshop on Machine Vision Applications, pp. 84–89.

[5] Freeman W & Roth M, (1995), "Orientation Histograms for Hand Gesture Recognition," IEEE Int. Workshop on Automatic Face and Gesture Recognition.

[6] Sawada H, Hashimoto S, and Matsushima T, (1998), "A Study of Gesture Recognition Based on Motion and Hand Figure Primitives and Its Application to Sign Language Recognition," vol. 39, no. 5, 1325-1333.

[7] Starner, T. and Pentland, A., (1995), "Real-time American Sign Language Recognition from Video Using Hidden Markov Models," 1995. Proc. of International Symposium on Computer Vision, vol. 21, no. 23, 265-270.

[8] Yoshino L et al, (1996), "Recognition of Japanese Sign Language from Image Sequence Using Color Combination," Proc. 3rd Int. Conf. Image Processing, 16-19.

[9] Ong SCW et al, (2005), "Automatic Sign language Analysis: A Survey and the Future beyond Lexical Meaning," IEEE Trans. PAMI, Vol. 27, No. 6, 873-891.

[10] Yamada Y, Matsuo T, Shimada N, and Shirai Y, (2009), Hand Detection and Hand Shape Classification Based on Appearance Learning for Sign Language Recognition, MIRU2009, IS1-37. (in Japanese)

[11] Tanaka S, Umeda Ki, (2001), Operating a Mobile Robot by Gesture Recognition, The transactions of the Institute of Electrical Engineers of Japan. C, A Publication of Electronics, Information and System Society, vol. 121, no. 9, pp. 1457-1463. (in Japanese)

[12] Fujimoto K, Matsuo T, Shimada N, and Shirai Y, (2010), "High Speed 3-D Hand Shape Measurement by Tree-based Learning Contour Features, Proc. of Meeting on Image Recognition and Understanding," MIRU2010, IS3-64. (in Japanese)

[13] Lee SU and Cohen I, (2004), "3D Hand Reconstruction from a Monocular View," Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04), Vol. 3, 310-313.

[14] Takeda Y, Terabayashi K, Asano H, and Umeda K, (2011), "Finger Direction Recognition in 3D Utilizing a Range and Image Sensor," Proc. Of Summer Seminar of Japan Soc. of Precision Engineering Technical Committee on Industrial Application of Image Processing, 75-76, (in Japanese)

[15] Ohkubo Y, Okada K, Inamura T, and Inaba M, (2005), "Finger Recognition for Target Indication to Robots in Daily Life (Humanoid 3, Mega-Integration in Robotics and Mechatronics to Assist Our Daily Lives)," Proc. of 2004 JSME Conference on Robotics and Mechatronics (ROBOMEC'04), 2A1-S-049.

[16] Rehg JM & Kanade T, (1994), "Visual Tracking of High DOF Articulated Structures: an Application to Human Hand Tracking," In Third European Conference on Computer Vision, 35-46.

[17] Nickel K & Stiefelhagen R (2007), "Visual Recognition of Pointing Gestures for Human-robot Interaction," Image and Vision Computing, vol. 25, no. 12, 1875-1884.

[18] Shotton J, Fitzgibbon A, Cook M, Sharp T, Finocchio M, Moore R, and Blake A (2011), "Real-time Human Pose Recognition in Parts from Single Depth Images," In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference, 1297-1304.

[19] Ikai T, Ohka M, and Yussof H (2012), "Behavior Control of Robot by Human Finger Direction," Engineering Procedia, 41, 784-791.

[20] Bradski G (2011), "Open Source Computer Vision Library," http://www.intel.com/research/mrl/research/opencv.

[21] Abdullah SC, Ikai T, Dosho Y, Yussof H and Ohka M (2011), "Edge Extraction Using Image and Three-axis Tactile Data," International Journal on Smart Sensing and Intelligent Systems, Vol. 4, No. 3, 508-526.

[22] Hashem, HF (2009), "Adaptive Technique for Human Face Detection Using HSV Color Space and Neural Networks," National Radio Science Conference, 2009, vol. 1, no. 7, 17-19

[23] Brooks, AG & Breazeal, C (2006), "Working with Robots and Objects: Revisiting Deictic Reference for Achieving Spatial Common Ground," In Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot Interaction, 297-304.

[24] Manders C, Farbiz F, Chong JH, Tang KY, Chua GG, Loke MH, and Yuan ML (2008), "Robust Hand Tracking Using a Skin Tone and Depth Joint Probability Model," 8th IEEE International Conference on Automatic Face & Gesture Recognition, Vol. 1, No. 6, 17-19

[25] Kim S, Sekiyama K, and Fukuda T, (2008)"Pattern Adaptive and Finger Image-guided Keypad Interface for In-vehicle Information Systems," International Journal on Smart Sensing and Intelligent Systems, Vol. 1, No. 3, 572-591.