https://doi.org/10.20965/jrm.2018.p0187

Paper:

Image Mosaicing Using Multi-Modal Images for Generation of Tomato Growth State Map

Takuya Fujinaga, Shinsuke Yasukawa, Binghe Li, and Kazuo Ishii

Kyushu Institute of Technology
2-4 Hibikino, Wakamatsu-ku, Fukuoka 808-0196, Japan
E-mail: fujinaga-takuya@edu.brain.kyutech.ac.jp
[Received October 3, 2017; accepted March 5, 2018]

Due to the aging and decreasing the number of workers in agriculture, the introduction of automation and precision is needed. Focusing on tomatoes, which is one of the major types of vegetables, we are engaged in the research and development of a robot that can harvest the tomatoes and manage the growth state of tomatoes. For the robot to automatically harvest tomatoes, it must be able to automatically detect harvestable tomatoes positions, and plan the harvesting motions. Furthermore, it is necessary to grasp the positions and maturity of tomatoes in the greenhouse, and to estimate their yield and harvesting period so that the robot and workers can manage the tomatoes. The purpose of this study is to generate a tomato growth state map of a cultivation lane, which consists of a row of tomatoes, aimed at achieving the automatic harvesting and the management of tomatoes in a tomato greenhouse equipped with production facilities. Information such as the positions and maturity of the tomatoes is attached to the map. As the first stage, this paper proposes a method of generating a greenhouse map (a wide-area mosaic image of a tomato cultivation lane). Using the infrared image eases a correspondence point problem of feature points when the mosaic image is generated. Distance information is used to eliminate the cultivation lane behind the targeted one as well as the background scenery, allowing the robot to focus on only those tomatoes in the targeted cultivation lane. To verify the validity of the proposed method, 70 images captured in a greenhouse were used to generate a single mosaic image from which tomatoes were detected by visual inspection.

Keywords: agriculture robot, tomato harvesting robot, infrared image, depth image, image mosaicing

1. Introduction

The number of agricultural workers in Japan is decreasing yearly, while their average age is increasing. According to statistical data for the five-year period from 2012 to 2016, the population of agricultural workers decreased by approximately 590,000, while the average age increased by approximately one year [a]. Furthermore, the food self-sufficiency rate in Japan is among the lowest among the major industrialized countries. To resolve these problems, methods to increase efficiency and precision are being adopted, such as high-yielding cultivation and computer-based environmental control. In addition, the use of robotic technology in agriculture is being considered to achieve automation [1, 2]. Studies on the automatic harvesting of cucumbers [3], asparagus [4], and green peppers [5] as well as studies those on the vision systems necessary for harvesting robots [6, 7] have been carried out.

In this study, we deal with tomatoes, which is one of the major types of vegetables. In comparison to cabbages or carrots, which also are major vegetables, tomatoes require more than five times the labor time per 1000 m^2 [b]. In particular, the ratios of harvesting and management (monitoring, training, pinching) are high in the labor hours of tomato cultivation. This is because the harvesting period for each tomato is different, and the process leading to harvest differs for individual tomatoes. Thus, a tomato greenhouse (large-scale greenhouse) that possesses environmental control functions and well-equipped production facilities is necessary to produce tomato efficiently and consistently

Hibikinada Greenhouse, Inc., with an interior area of 8.5 hectares, is an example of such a large-scale greenhouse, and is the site where we conduct harvesting experiments and investigate specific needs. The basic layout of the tomato greenhouse is shown in **Fig. 1**. The tomatoes are cultivated in a single row parallel to the working lane. The ripe tomatoes are arranged in positions 80 to 120 cm above the ground, which allows the workers to harvest them with ease. At Hibikinada Greenhouse, there are several tens of cultivation lanes, each of which is approximately 70 m long. In the working lane, pipes are installed through which warm water flows to keep the interior temperature constant. The pipes also serve as the rails for carts used by the workers during harvesting.

However, the working environment in the greenhouse is hot and humid and very uncomfortable for workers. Studies on tomato harvesting robots [8–10] have been conducted to reduce the burden on workers. While many studies on automating agricultural work have focused on the harvesting of crops, the workers seek the automation

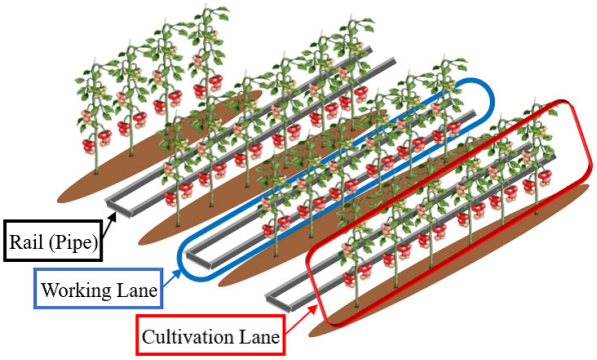


Fig. 1. Layout of tomato greenhouse.

Fig. 2. Tomato growth state map.

of not only the harvesting but also the management of agricultural produce as well.

We are engaged in the research and development of robot that not only harvest crops in large-scale greenhouses but also manage the growth state of crops [11]. In this paper, management refers to the monitoring of crops. The purpose of our study is to develop a system that automatically harvest and manage tomatoes. To achieve automatic harvesting, it is necessary to automatically detect the positions of harvestable tomatoes and automatically plan the harvesting motions of the robot. To develop a management system, it is necessary to automatically detect the tomatoes and judge the maturity of the tomatoes in the greenhouse. In addition, the yield and harvesting period must be predicted. In this paper, we propose as the first stage a method to generate a greenhouse map (a wide-area mosaic image of the tomato cultivation lane). It is assumed that the greenhouse map enables workers to monitor and grasp the growth state of tomatoes.

As **Fig. 1** shows, since there are multiple cultivation lanes in the greenhouse, many tomatoes are present in the background when images of the target lane are captured by a camera. Therefore, it is necessary to eliminate the cultivation lanes lying behind the one we wish to capture as well as the other background part so that the tomato-harvesting robot and the worker monitoring the greenhouse map can focus on the targeted tomatoes. This will prevent the robot making erroneous detections of tomatoes in lanes behind and allow the worker to concentrate on grasping the growth state of the targeted tomatoes.

Although there has been a study to detect tomatoes in the front lane while eliminating the background [12], it employs images captured from a position relatively close to the tomatoes (20 cm) and is usable only in restricted environments. Furthermore, this study [12] does not allow measurement of the three-dimensional positions of tomatoes. The three-dimensional positions of tomatoes need when the robot harvest tomato.

The rails installed in Hibikinada Greenhouse are approximately 65 cm wide, so the tomatoes to be harvested lie at a distance between approximately 65 cm and 1 m from the camera. In the method proposed in this study, to determine the three-dimensional positions of the tomatoes, we employ a time-of-flight (ToF) based Kinect camera (hereafter, Kinect v2), which acquires depth images that provide information on the distance from the camera to an object. In addition to depth images, Kinect v2 can also capture red-green-blue (RGB) and infrared images.

To eliminate the cultivation lane behind and generate a mosaic image consisting of only the targeted cultivation lane, infrared, depth, and RGB images are used to generate the greenhouse map. In this proposed method, a correspondence point problem of feature points eases.

The final goal is to generate a tomato growth state map, as shown in **Fig. 2**, and implement a system to automatic harvest and manage tomatoes by using a robot (the cultivation lane behind and background scenery have not been eliminated in **Fig. 2**). The tomato growth state map provides information on the position and maturity of tomatoes.

In this paper, we propose a method to generate the greenhouse map, which will be used as the basic image for the tomato growth state map. To verify the validity of the proposed method, we generated a single greenhouse map from 70 images captured inside the Hibikinada Greenhouse, and used it to detect tomatoes by visual inspection. The results are presented.

2. Verification of Method to Generate Greenhouse Map

2.1. Flow of Automatic Harvesting of Tomatoes and Method of Generating Greenhouse Map

The robot selects the rail to be targeted and moves on the rails (**Fig. 3** (A)). As it moves, the robot captures images of the cultivation lane. As shown in **Fig. 3(b)**, the robot captures images at constant intervals t_x, while the distance between the camera and tomatoes is t_z. The image capture positions are measured using an encoder mounted on the moving mechanism. When the robot reaches the end of the rails, it generates the tomato growth state map (**Fig. 3** (B)). Then, it detects the harvestable tomatoes from the map and shifts to begin harvesting motions (**Fig. 3** (C)).

In this study, a greenhouse map is generated from images acquired by the robot on the rails. While it is possible to employ a fisheye lens or omnidirectional sensor to generate a greenhouse map, a system using a special lens or sensor will be relatively costly. In addition, such sys-

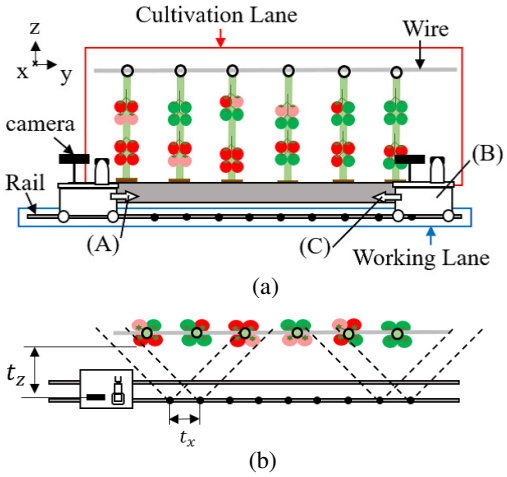


Fig. 3. Flow of automatic harvesting of tomatoes: (a) front view, (b) top view.

tems capture a wide range in a single image, so that the resolution tends to be low.

However, image mosaicing is a method that can be used to produce high-resolution images with a wide visual field using only an ordinary camera. In the proposed method, a mosaic image is produced by stitching multiple images to generate the greenhouse map.

Digital maps and aerial photographs are examples of mosaic images. In addition, mosaic images are used in geomorphological studies [13] and biological studies in marine industries [14]. There have also been studies on attaching information to mosaic images [15] as well as studies on monitoring systems based on mosaic images [16].

The goal of this work is to generate a tomato growth state map which consists of a mosaic image (greenhouse map) to which information is attached, such as the positions and maturity of tomatoes, harvesting period, and so forth.

2.2. Generation of a Mosaic Image

This section describes the method of generating a mosaic image. As an example, we consider the generation of a single mosaic image from two images. Two images captured from adjacent positions are prepared. The two images contain part that overlap one another. Feature points are extracted in the two images, and then the feature points are corresponded based on their features. The two images are stitched using the corresponding feature points to generate the mosaic image.

The mosaic image is generated by using a homography matrix computed from the correspondence pairs of feature points. The homography matrix consists of eight parameters, and requires four or more correspondence pairs. A slight change in a single parameter can produce great distortions when the mosaic image is generated. Therefore, studies have been carried out to automatically select the transformation model for the homography matrix based on the correspondence pairs to stably generate a mosaic image [17]. In the proposed method, we employ the moving distance by the robot to determine the transformation model of the homography matrix. Details are provided in Section 4.4.

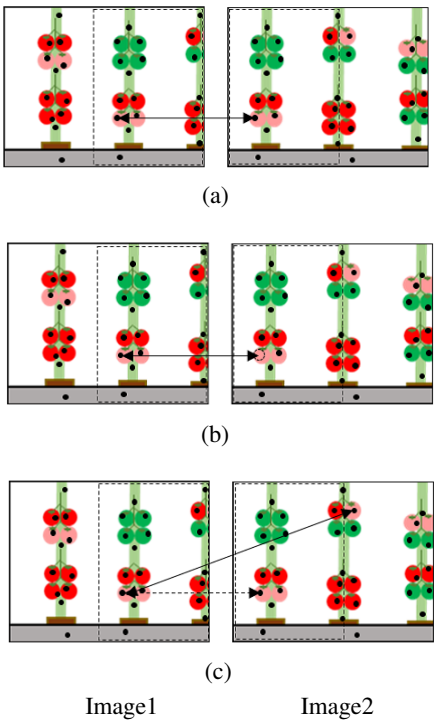


Fig. 4. Repeatability and distinctiveness: (a) repeatability and distinctiveness both satisfied, (b) repeatability not satisfied, (c) distinctiveness not satisfied.

2.3. Conditions for Generating a Mosaic Image

The following three conditions are considered necessary to generate a mosaic image:

1. repeatability of feature points
2. distinctiveness of feature points
3. assume that the captured image is as a planar surface.

2.3.1. Conditions 1 and 2

In this subsection, we discuss the repeatability and distinctiveness of feature points. Repeatability means that the feature point is always extracted as a feature point; distinctiveness means that the features of a feature point can be distinguished from those of another feature point.

The images shown in **Fig. 4** explain conditions 1 and 2. The segments surrounded by broken lines are the overlapping part in the images captured at adjacent positions, while the dots (•) indicate the extracted feature points. When repeatability and distinctiveness are both satisfied (**Fig. 4(a)**), feature points are extracted in the same (corresponding) positions in images 1 and 2, and their features are distinguishable from those of other feature points. Thus, the feature points of images 1 and 2 can be corresponded correctly.

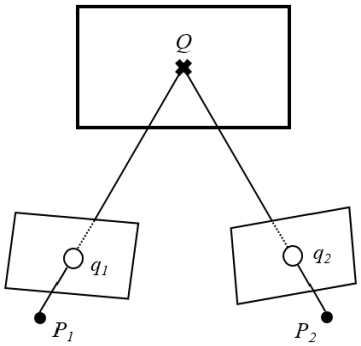


Fig. 5. Planar projection.

When the feature points do not have repeatability (**Fig. 4(b)**), the feature point extracted in image 1 cannot be corresponded with another at the corresponding position (indicated by a circle in image 2). When the feature points do not have distinctiveness (**Fig. 4(c)**), the feature points extracted in the same positions in images 1 and 2 possess features similar to those of other feature points, so that correspondence cannot be carried out correctly.

Since corresponding feature points are searched in the overlapping sections of the two images captured from adjacent positions when a mosaic image is generated, the correspondence point problem must be solved. In general, strategies are employed to reduce false correspondences by limiting the feature points to be corresponded or limiting the area searched for corresponding points [18]. In this study, we employ the latter, i.e., limiting the search area for corresponding points. Details are provided in Section 4.3.

2.3.2. Condition 3

In image mosaicing, the geometric characteristic of the captured objects must be limited. When the targeted scene is that of a distant view or, if it lies closer, consists of a flat plane, such as a building or wall, it can be assumed to form a single plane. When a point Q in three-dimensional space is observed from points P_1 and P_2, as shown in **Fig. 5**, and all of the observed points lie on some plane in three-dimensional space, then the coordinates in the respective images, q_1 and q_2, are known to have a linear relation [19].

In other words, point $q_1 = (x_1, y_1, 1)$, expressed in homogenous coordinates, has a corresponding point $q_2 = (x_2, y_2, 1)$, whose relationship is defined by Eq. (1), which is known as a homography. In Eq. (1), h_0, h_1, h_2, h_3, h_4, h_5, h_6, and h_7 are the eight parameters of the homography matrix mentioned in Section 2.2.

$$\begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix} = \begin{pmatrix} h_0 & h_1 & h_2 \\ h_3 & h_4 & h_5 \\ h_6 & h_7 & 1 \end{pmatrix} \begin{pmatrix} x_2 \\ y_2 \\ 1 \end{pmatrix} \quad \ldots\ldots \quad (1)$$

Previous studies have deal with distant views or planar scenes [13–17, 20]. For example, consider the image-capture environment shown in **Fig. 6(a)**. Assume that images are captured at two adjacent points (points P_1 and P_2) of an object lying at a distance d [m] from the position captured. If the target scene is at a large distance ($d = 1000$ m), a mosaic image such as that shown in **Fig. 6(b)** is generated [20]. In this case, condition 3 is satisfied.

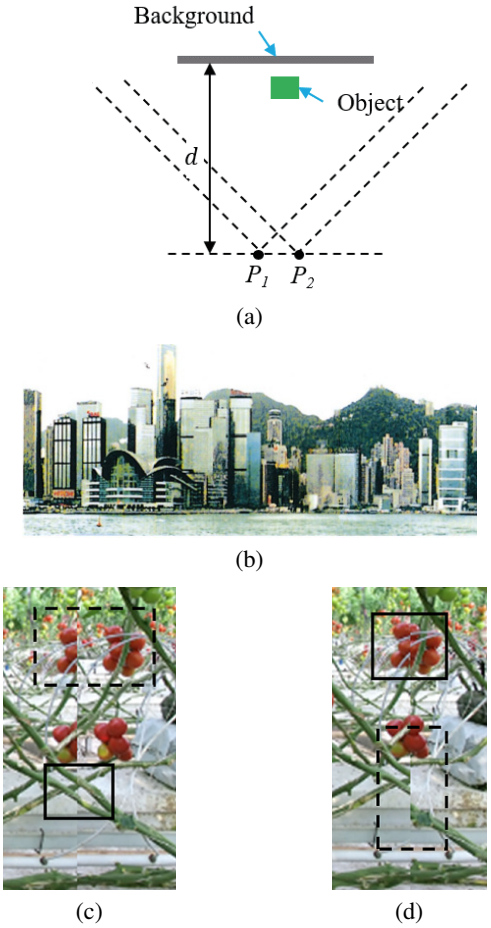


Fig. 6. Image-capture environment for condition 3 and mosaic images of different scenes: (a) image-capture environment, (b) mosaic image when captured object is distant ($d = 1000$ m) [20], (c) seam in mosaic image for $d = 0.65$ m when attention is focused on object and (d) when attention is focused on background.

On the other hand, if the targeted scene is at a close distance ($d = 0.65$ m), as in **Figs. 6(c)** and **(d)**, and the camera cannot decide on which plane to focus, condition 3 is not satisfied. In this case, different mosaic images are generated depending on the object on which attention is focused. Thus, the image shown in **Fig. 6(c)** is generated when attention is focused on the object in **Fig. 6(a)**, while the image shown in **Fig. 6(d)** is generated when attention is focused on the background.

No gaps exist at the seam between images in the areas surrounded by solid lines in **Figs. 6(c)** and **(d)**. However, there are gaps at the seam between images in the areas surrounded by broken lines. If distance d is relatively small and cannot be assumed to be a plane, the mosaic image differs depending on the object on which attention is focused.

(a) (b)

Fig. 7. Captured images of cultivation lane: (a) RGB image, (b) infrared image.

3. Use of Infrared Image

An RGB image of a cultivation lane captured from a position above the rails is shown in **Fig. 7(a)**. Because the fruits, stems and leaves grow close together in the greenhouse, and all plants have similar characteristics, it is easy for false correspondences to occur. The working lane has a width of approximately 65 cm, the space in which the robot can move is limited, and the distance between the camera and the harvestable tomatoes is 0.65 m to 1.0 m, which is quite close. The following problems can arise in attempts to generate a mosaic image from RGB images of a cultivation lane.

1. Because objects with similar characteristics are located close together, false correspondences can occur.

2. The capturing environment makes it difficult to decide the plane on which attention should be focused.

To resolve problems 1 and 2, we employ time-of-flight (ToF) infrared images in the proposed method. An infrared image of a cultivation lane captured from a position above the rails is shown in **Fig. 7(b)**. In the ToF method, the time required for a laser beam projected by the device to be reflected by a target object and return is measured. In a ToF-based infrared image, the intensity of the reflected infrared light is measured. The light reflected by a distant object has a low intensity, while that from the tomatoes in the lane in front, which are close, has a high intensity.

As seen in **Fig. 7**, the number of objects (fruits, stems and leaves) in the infrared image is low. Since Kinect v2 cannot capture infrared images of objects lying at a distance of 8 m or more, those objects can be ignored. Furthermore, an infrared image can be used to generate a depth image possessing depth information. Thus, by using two types of images, it is possible to eliminate the cultivation lane behind as well as the general background and capture an image of the tomatoes in the front lane.

We can expect to reduce the number of false correspondences of feature points by limiting the depth of the capture range to the cultivation lane, in which the fruits, stems, and leaves are closely located. By focusing attention on only those tomatoes in the front lane, it is possible to treat a given depth as a plane. In proposed method, infrared images and depth images are used to solve problems 1 and 2 to generate mosaic images.

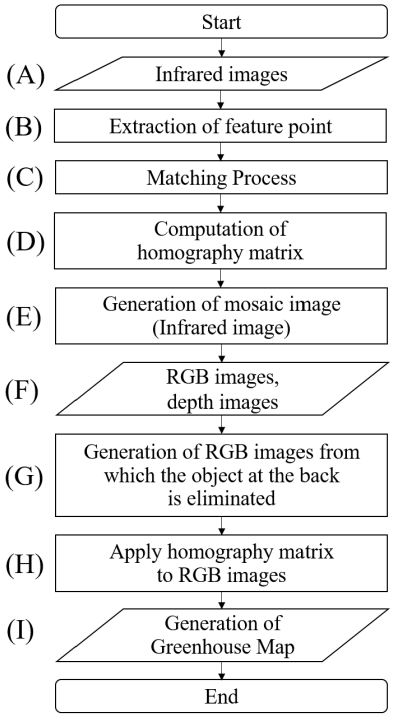


Fig. 8. Algorithm for generating greenhouse map.

In addition, real-time processing is required to operate the robot in a real environment. The use of infrared images to generate a mosaic image is a valid method considering that the robot must process this information. While an RGB image carries 24 bits per pixel, since R, G and B are each specified by 8 bits, an infrared image consists of 16 bits per pixel. In other words, an infrared image carries two-thirds the information of an RGB image.

4. Method of Generating a Greenhouse Map Using Infrared Images

Figure 8 shows the algorithm to generate a greenhouse map using infrared images. Using infrared images as input (**Fig. 8** (A)), feature points are extracted (**Fig. 8** (B)). In this paper, we used speed-up robust features (SURF) [21]. After limiting the search area for feature points to reduce false correspondences, the feature points are matched (**Fig. 8** (C)). The matching process is described in detail in Sections 4.2 and 4.3. The correspondence pairs are used to compute the homography matrix (**Fig. 8** (D)), which is used to generate a mosaic image from the infrared images (**Fig. 8** (E)).

Next, the RGB and depth images are used as inputs to focus only the tomatoes in the front lane (**Fig. 8** (F)). The depth images are used to extract only those tomatoes lying in the harvestable area (up to 1 m from the camera) to construct an RGB image (**Fig. 8** (G)). The homography

matrix computed from the infrared images is then applied to the RGB images (**Fig. 8** (H)) to generate a greenhouse map (**Fig. 8** (I)).

4.1. Conditions for Capturing Images of Cultivation Lane and Camera Parameters

As shown in **Fig. 3(a)**, the robot captures images of the cultivation lane with a mounted camera (Kinect v2) as it moves on the rails. The preconditions for image capture are summarized as follows:

1. Images are captured in sequence at constant intervals of t_x.
2. The distance between the camera and the tomatoes in the front lane is constant, at t_z.
3. The camera is fixed to the robot and is unable to turn.

The rotation matrix $\boldsymbol{R}$ and translation vector $\boldsymbol{t}$, which are external parameters of the camera, are given by Eqs. (2) and (3), respectively. Here, R_{11}, R_{12}, R_{13}, R_{21}, R_{22}, R_{23}, R_{31}, R_{32} and R_{33} represent the elements of the rotation matrix, and t_x, t_y, and t_z represent the distances of translation, in units of millimeters.

In proposed method, it is assumed that the robot moves in intervals of 300 mm (t_x) on the rails to capture images and that the distance (t_z) between the camera and the tomatoes on the front lane is 650 mm. To verify the validity of the proposed approach, the robot is moved in manually for distances of 300 mm on the rails to capture images of the cultivation lane.

MATLAB Camera Calibrator [22] is used to calibrate the camera. The images of 15 calibration boards are captured and used as inputs. The internal parameters estimated by Camera Calibrator are given in Eq. (4), where f_x and f_y are the focal length in pixels, c_x and c_y are the optical centers in pixels, and s is the skew coefficient.

$$\boldsymbol{R} = \begin{pmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \ldots \quad (2)$$

$$\boldsymbol{t} = \begin{pmatrix} t_x & t_y & t_z \end{pmatrix} = \begin{pmatrix} 0 & 0 & 650 \end{pmatrix} \quad \ldots \ldots \quad (3)$$

$$\boldsymbol{K} = \begin{pmatrix} f_x & 0 & 0 \\ s & f_y & 0 \\ c_x & c_y & 1 \end{pmatrix} = \begin{pmatrix} 346.3 & 0 & 0 \\ 0 & 346.8 & 0 \\ 256.8 & 205.8 & 1 \end{pmatrix} \quad . \quad (4)$$

4.2. Basic Experiment to Determine the Search Area for Correspondence Points

As stated in Section 2.3.1, the correspondence point problem of feature points must be solved to generate a mosaic image of a greenhouse. In this study, we reduced the false correspondences of feature points by limiting the search area using the moving distance of the robot. In this section, we describe the preliminary experiment conducted to determine the range of the search area. The test board shown in **Fig. 9(a)** is captured by a camera at fixed intervals d_x, as shown in **Fig. 9(b)**. In the test board, the red area repeats at intervals of l.

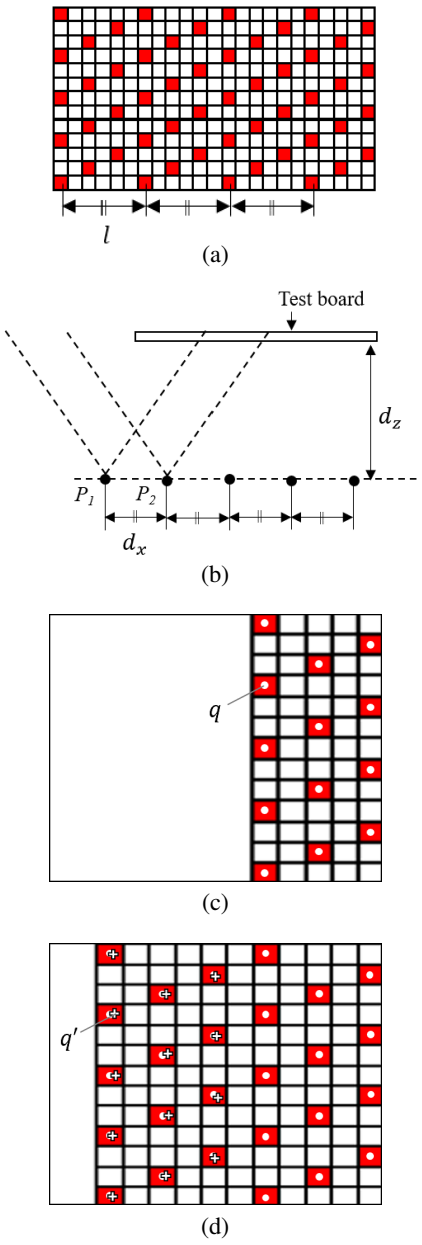
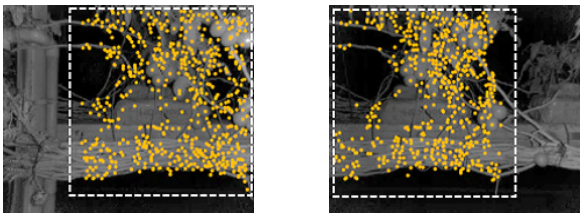


Fig. 9. (a) Test board used in preliminary experiment, (b) configuration of preliminary experiment, (c) positions of centroids (•) of markers in image captured from point P_1, (d) positions of centroids (•) and estimated points (+) of markers in image captured from point P_2.

The images capture at points P_1 and P_2 are shown in **Figs. 9(c)** and **(d)**, respectively. To compute the estimated point, the centroid q in **Fig. 9(c)** is transformed from the image coordinate system to the world coordinate system using the camera's internal and external parameters (Eqs. (4), (2), and (3), respectively). The relation between the image coordinate system (u, v) and the world coordinate system (X, Y, Z) is given by Eq. (5), where λ is the depth information in image coordinates.

$$\lambda \begin{pmatrix} u & v & 1 \end{pmatrix} = \begin{pmatrix} X & Y & Z & 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{R} \\ \boldsymbol{t} \end{pmatrix} \boldsymbol{K} \quad \ldots \quad (5)$$



(a) (b)

Fig. 10. Extracted feature points in infrared images: (a) infrared image 1, (b) infrared image 2.

Since the camera moved from point P_1 to P_2, d_x is added to the X element in the world coordinate system, and then the world coordinate system is transformed back to the image coordinate system using Eq. (5). The point of the estimated point is denoted q'. The same transformation is carried out for all centroids in **Fig. 9(c)**. In **Figs. 9(c)** and **(d)**, the dots (•) represent the centroids, and the plus signs (+) represent the estimated points.

Using the Kinect v2 camera mounted on the robot, images of the test board were captured from the five points indicated in **Fig. 9(b)**. The interval between points, d_x, was 300 mm; the distance from the camera to the test board, d_z, was 650 mm; and the horizontal distance between red sections, l, was 300 mm. The images were captured by manually moving the robot at fixed intervals ($d_x = 300$ mm). Note that d_x and d_z have the same values as the interval at which images were captured in the greenhouse, t_x, and the distance between the camera and the tomatoes on the front lane, t_z, respectively. From the five images, the errors of the estimated points against the centroids were determined. The closer the estimated point is to the centroid, the more accurate the calculation of estimated point is.

The centroid was found to be at most 10 pixel off in both the positive and negative x-axis directions, and at most 6 pixel off in the positive and negative y-axis directions. The standard deviations were 4.3 pixel and 2.0 pixel pixels in the x- and y-axis directions, respectively.

Based on these results, the search area range was confined to a rectangular area 20 pixel wide (horizontal) and 12 pixel deep (vertical), centered at the estimated point.

4.3. Matching Process Based on the Robot's Moving Distance

In this section, we describe the matching process based on the robot's moving distance, which is adopted to reduce false correspondences of feature points. First, the feature points are extracted from two infrared images which have overlapping part. The interval at which the two images are captured is constant, so that the overlapping area is known. Therefore, feature points are extracted from only the overlapping areas, as indicated by the area surrounded by broken lines in **Fig. 10**, to reduce false correspondences. The dots (•) in **Fig. 10** represent the extracted feature points.

Next, we focus on a feature point in infrared image 1 (left side) in **Fig. 11**. Based on this feature point and the robot's moving distance t_x, the estimated point is drawn on infrared image 2 (right side). The estimated point is calculated using Eq. (5). The robot's moving distance t_x is 300 mm.

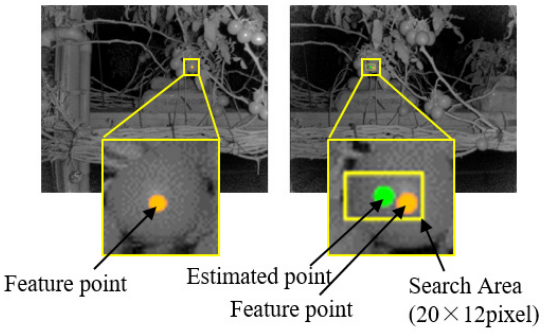


Fig. 11. Search area and feature points in search area.

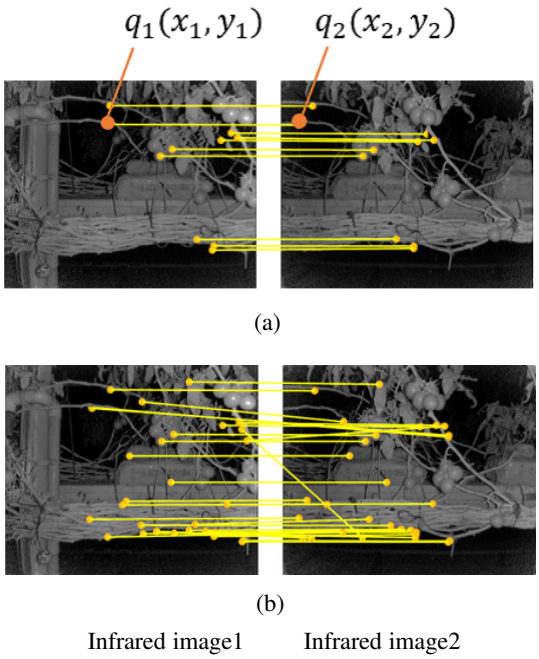


Fig. 12. Correspondences of feature points in two input images: (a) correspondence result of feature points when robot moving distance was used, (b) correspondence result of feature points when robot moving distance was not used.

Then, the features of a feature point located within the search area, which are determined as described in Section 4.2, centered at the estimated point in infrared image 2, are compared with those of the feature point in infrared image 1 to match those two points.

The same procedure was carried out for all feature points in infrared image 1. The matched (i.e., correspondence) results are shown in **Fig. 12(a)**. **Fig. 12(b)** shows the correspondence results when the robot's moving distance was not used. In addition, the correspondence results when the RGB images were used instead of infrared images are shown in **Fig. 13**. Note that the robot's moving distance was not used for matching in **Fig. 13**.

Fig. 13. Correspondence result of feature points using RGB images.

4.4. Computation of the Homography Matrix

The three preconditions for image capture were stated in Section 4.1. Following those preconditions, a translation model, with two parameters but without those for rotation and magnification/contraction, is applied to the homography matrix. This matrix is given by Eq. (6), where d_x and d_y are respectively the moving distances in the x- and y-axis directions. Here, d_x and d_y represent the moving distances (pixel) in the x- and y-axis directions within the two images, and they are determined from the corresponding feature points. For q_1 and q_2 in **Fig. 12(a)**, d_x and d_y are given by Eqs. (7) and (8), respectively.

$$\boldsymbol{H} = \begin{pmatrix} h_0 & h_1 & h_2 \\ h_3 & h_4 & h_5 \\ h_6 & h_7 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & d_x \\ 0 & 1 & d_y \\ 0 & 0 & 1 \end{pmatrix} \quad \ldots \quad (6)$$

$$d_x = x_1 - x_2 \quad \ldots \quad (7)$$

$$d_y = y_1 - y_2 \quad \ldots \quad (8)$$

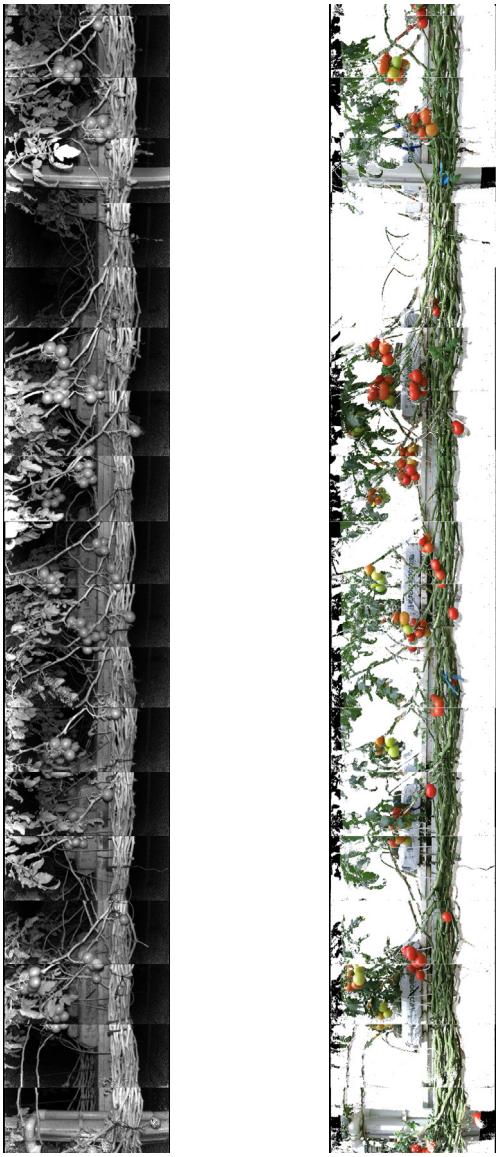
4.5. Generation of a Mosaic Image Based on Infrared Images and Greenhouse Map

The computed homography matrix is used to generate the mosaic image. The homography matrix is also applied to the RGB images to generate the greenhouse map. The RGB images in this case are those in which only the information of tomatoes in close range has been extracted based on information in the depth images. A single mosaic image was generated from 70 images captured from the actual environment. The mosaic image has a length of approximately 20 m.

In this study, we show one-quarter of this image. The mosaic images based on infrared images and RGB images (greenhouse map) are shown in **Figs. 14(a)** and **(b)**, respectively. The tomato growth state map is generated by attaching information, such as the positions and maturity of tomatoes, to the mosaic image of **Fig. 14(b)**. Tomatoes were detected by visual inspection. As a result, a total of 302 tomatoes, of which 184 were ripe tomatoes, were identified in the entire mosaic image.

4.6. Correspondence Results of Feature Points

In the correspondence pairs of feature points in **Fig. 12**, the coordinates of a feature point in infrared image 1 are denoted by $q_1(x_1, y_1)$, and those of the corresponding feature point in infrared image 2 are denoted by



(a) (b)

Fig. 14. One-quarter of mosaic image generated from 70 images: (a) mosaic image using infrared images, (b) mosaic image using RGB images.

$q_2(x_2, y_2)$. The range of correct correspondence was determined from the robot's moving distance. The criterion for a correct correspondence in this study is given as follows:

$$150 \text{ pixel} < x_1 - x_2 < 170 \text{ pixel}$$
$$-6 \text{ pixel} < y_1 - y_2 < 6 \text{ pixel}$$

The correspondence is correct when the above conditions are satisfied and incorrect when they are not. The true value of the displacement in the x-axis direction was determined from the robot's moving distance (300 mm) to be 160 pixel. The true value of the displacement in the y-axis direction is 0 pixel since the camera is fixed on the robot. However, due to the robot's moving distance and slight differences in the level of the rails (approximately 1 to 3 mm), the range of the search area, found in the

Table 1. Correspondence result of feature points.

	Infrared image		RGB image
Pair No.	result 1	result 2	result 3
1	$\frac{10}{10}$ (100%)	$\frac{5}{39}$ (12.8%)	$\frac{0}{958}$ (0.0%)
2	$\frac{9}{9}$ (100%)	$\frac{7}{35}$ (20.0%)	$\frac{0}{1026}$ (0.0%)
3	$\frac{10}{10}$ (100%)	$\frac{4}{35}$ (11.4%)	$\frac{0}{1226}$ (0.0%)
4	$\frac{5}{5}$ (100%)	$\frac{2}{46}$ (4.3%)	$\frac{0}{1119}$ (0.0%)
5	$\frac{11}{11}$ (100%)	$\frac{5}{35}$ (14.3%)	$\frac{2}{1110}$ (0.2%)
6	$\frac{10}{10}$ (100%)	$\frac{5}{44}$ (11.4%)	$\frac{1}{957}$ (0.1%)
7	$\frac{10}{10}$ (100%)	$\frac{6}{44}$ (13.3%)	$\frac{1}{1028}$ (0.1%)
8	$\frac{7}{7}$ (100%)	$\frac{4}{37}$ (10.8%)	$\frac{1}{1135}$ (0.1%)
9	$\frac{6}{6}$ (100%)	$\frac{3}{48}$ (6.3%)	$\frac{1}{1272}$ (0.1%)
10	$\frac{6}{6}$ (100%)	$\frac{6}{48}$ (12.5%)	$\frac{1}{1192}$ (0.1%)

preliminary experiment of Section 4.2, was used as the allowable error range.

The corresponding feature points were confirmed using ten pairs of images, each pair consisting of two images which were captured from adjacent positions and contained overlapping part. The correspondence results were verified for three cases: those obtained from infrared images using the robot's moving distance (result 1), those obtained without using the robot's moving distance (result 2), and those obtained from RGB images without the robot's moving distance (result 3). The results are presented in **Table 1**. In each case, the numerator and denominator represent the numbers of correct correspondences and total correspondences, respectively, while the table in parenthesis give the percentages of correct correspondences.

5. Discussion

5.1. Evaluation of Matching Process

In the proposed method, the estimated point is computed from a given feature point, and a search is made for the corresponding feature point within a search area centered at the estimated point. As seen in **Fig. 12**, the number of grossly false correspondences can be reduced by using the robot's moving distance to correspond feature points.

The percentage of correct correspondences was presented in Section 4.6 to evaluate the results. First, we use **Table 1** to compare the infrared images (result 2) and RGB images (result 3) for which the search area was not limited. Although RGB images yield a higher number of correspondences than infrared images, the percentage of correct correspondences is extremely low. The RGB image contains more objects, such as the fruits, stems, and leaves, compared to the infrared image. Since these objects have similar features, the feature point in the second image that corresponds to a given feature point in the first image was not identified correctly, which resulted in the high number of false correspondences. From this result, we can state that RGB images of the cultivation lane are not satisfactory for distinctiveness of feature points (condition 2).

(a) (b)

Fig. 15. Discontinuities in seam between images: (a) without discontinuity, (b) with discontinuity.

As the proposed method applies a translation model of the homography matrix, at least one correct correspondence must be found. However, computation of the homography matrix is also affected by correspondence pairs other than the correct pairs. Thus, a slight change in one of the eight parameters of the homography matrix can produce large distortions when a mosaic image is generated. Thus, false correspondences are present in result 2, showing that the use of only infrared images is not sufficient to generate the desired mosaic image. False correspondences were found in none of the pairs in result 1. Thus, a mosaic image such as the one shown in **Fig. 14** can be generated by using infrared images and limiting the search area for feature points. It is important to use the correct correspondence pairs to generate a mosaic image.

5.2. Gaps Between Images in a Mosaic Image

Parts of the generated mosaic image are shown in **Fig. 15**. The image in **Fig. 15(a)** displays no gaps in the seam between two images, but the image in **Fig. 15(b)** displays a gap.

The estimated point is computed based on the robot's moving distance, d_x (300 mm), and the distance from the camera to the tomatoes on the front lane, d_z (650 mm). A gap occurs in the seam between two images when a portion that does not coincide with the distance d_z. The information of the depth image at the seam in **Fig. 15(a)** is given as 650 mm $\pm$ 20 mm, but that in **Fig. 15(b)** is given as 750 mm $\pm$ 20 mm.

In the proposed method, the mosaic image is generated from images captured from points at constant intervals. However, for images in which tomatoes are clustered together at the center, we think that it is possible to prevent discontinuities from occurring between images by reducing the overlapped areas.

5.3. Generating a Mosaic Image Considering the Robot's Moving Distance

To verify the proposed method, the robot was manually moved in steps of 300 mm to capture images in this study.

To construct a system to automatic harvesting and management of tomatoes, it is necessary to generate a greenhouse map using the robot's moving distance. The moving mechanism has an encoder, the information of which can be used to measure the moving distance.

A critical issue in this regard is that the estimated point is determined from the moving distance. The estimated point is an important parameter affecting the accuracy of the correspondence, since it is used to determine the search area. If the wheels slip when the robot moves on the rails, this will affect the moving distance as well as the correspondence of feature points. Thus, it is necessary to take into account any error in the moving distance when implementing the proposed system using the robot.

6. Conclusion

In this paper, we proposed a method to generate a mosaic image based on infrared images. The use of infrared images satisfies the conditions necessary to generate a mosaic image. It is also possible to obtain an image of only the target cultivation lane by using infrared images and depth images. By applying the homography matrix computed from infrared images to RGB images, it is possible to generate an RGB mosaic image. We were able to identify a total of 302 tomatoes of which 184 were ripe from the mosaic image produced from 70 images.

To follow up this study, we plan to automatically detect and mark tomatoes in the greenhouse map, making it possible to estimate the yield and harvest time.

References:

[1] J. Sato, "Farming Robots," J. Rob. Mech., Vol.9, No.4, pp. 287-292, 1997.

[2] N. Kondo and M. Monta, "Fruit Harvesting Robotics," J. Rob. Mech., Vol.11, No.4, pp. 287-292, 1999.

[3] S. Arima and N. Kondo, "Cucumber Harvesting Robot and Plant Training System," J. Rob. Mech., Vol.11, No.3, pp. 208-212, 1999.

[4] S. Bachche and K. Oka, "Design, Modeling and Performance Testing of End-Effector for Sweet Pepper Harvesting Robot Hand," J. Rob. Mech., Vol.25, No.4, pp. 705-717, 2013.

[5] N. Irie, N. Taguchi, T. Horie, and T. Ishimatsu, "Development of Asparagus Harvester Coordinated with 3-D Vision Sensor," J. Rob. Mech., Vol.21, No.5, pp. 583-589, 2009.

[6] N. Noguchi, J. F. Reid, Q. Zhang, L. Tian, and A. C. Hansen, "Vision Intelligence for Mobile Agro-Robotics System," J. Rob. Mech., Vol.11, No.3, pp. 193-199, 1999.

[7] M. Monta, N. Kondo, S. Arima, and K. Namba, "Robotic Vision for Bioproduction Systems," J. Rob. Mech., Vol.15, No.3, pp. 341-348, 2003.

[8] N. Kondo, K. Yamamoto, H. Shimizu, and K. Yata, "A machine vision system for tomato cluster harvesting robot," Engineering in Agriculture, Environment and Food, Vol.2, No.2, pp. 60-65, 2009.

[9] H. Yaguchi, K. Nagahama, T. Hasegawa, and M. Inaba, "Development of an autonomous tomato harvesting robot with rotational plucking gripper," 2016 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS), pp. 652-657, 2016.

[10] W. Lili, Z. Bo, F. Jinwei, H. Xiaoan, W. Shu, L. Yashuo, Z. Qiangbing, and W. Chongfeng, "Development of a tomato harvesting robot used in greenhouse," Int. J. Agric. and Bio. Eng., Vol.10, No.4, pp. 140-149, 2017.

[11] S. Yasukawa, B. Li, T. Sonoda, and K. Ishii, "Development of a Tomato Harvesting Robot," The 2017 Int. Conf. on Artificial Life and Robotics (ICAROB 2017), pp. 408-411, 2017.

[12] R. F. Teimourlou, A. Arefi, and A. M. Motlagh, "A Machine Vision System for the Real-Time Harvesting of Ripe Tomato," J. Agricultural Machinery Science, Vol.7, No.2, pp. 159-165, 2011.

[13] G. G. Michael, S. H. G. Walter, T. Kneissl, W. Zuschneid, C. Gross, P. C. McGuire, A. Dumke, B. Schreiner, S. Gasselt, K. Gwinner, and R. Jaumann, "Systematic processing of mars express HRSC panchromatic and colour image mosaics: Image equalisation using an external brightness reference," Planetary and Space Science, Vol.121, pp. 18-26, 2016.

[14] K. Jerosch, A. Ludtke, M. Schluter, and G. T. Ioannidis, "Automatic content-based analysis of georeferenced image data: Detection of Beggiatoa mats in seafloor video mosaics from the Hakon Mosby Mud Volcano," Computers and Geosciences, Vol.33, pp. 202-218, 2007.

[15] E. H. Helmer, T. S. Ruzycki, J. M. Wunderle Jr, S. Vogesser, B. Ruefenacht, C. Kwit, T. J. Brandeis, and D. N. Ewert, "Mapping tropical dry forest height, foliage height profiles and disturbance type and age with a time series of cloud-cleared Landsat and ALI image mosaics to characterize avian habitat," Remote Sensing of Environment, Vol.114, pp. 2457-2473, 2010.

[16] T. Suzuki, Y. Amano, T. Hashizume, S. Suzuki, and A. Yamada, "Generation of Large Mosaic Images for Vegetation Monitoring a Using Small Unmanned Aerial Vehicle," J. Rob. Mech., Vol.22, No.2, pp. 212-220, 2010.

[17] Y. Kanazawa and K. Kanatani, "Stabilizing Image Mosaicing by Model Selection," 3D Structure from Images – SMILE 2000, pp. 35-51, 200.

[18] D. Marr, "Vision," San Francisco: W. H. Freeman and Company, 1982.

[19] O. Faugeras, "Three-Dimensional Computer Vision: A Geometric Viewpoint," MIT Press, 1993.

[20] P. Bao and D. Xu, "Complex wavelet-based image mosaic using edge-preserving visual perception modeling," Computer and Graphics, Vol.23, pp. 309-321, 1999.

[21] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF:Speeded Up Robust Features," Computer Vision and Image Understanding (CVIU), Vol.110, No.3, pp. 346-359, 2008.

[22] Z. Zhang, "A Flexible New Technique for Camera Calibration," IEEE Trans. on Pattern Analysis and Machine Intelligence. Vol.22, No.11, pp. 1330-1334, 2000.

Supporting Online Materials:

[a] Ministry of Agriculture, Forestry and Fisheries
http://www.maff.go.jp/j/tokei/sihyo/data/08.html
[Accessed January 30, 2018]

[b] Survey on Agricultural Management
http://www.e-stat.go.jp/SG1/estat/List.do?lid=000001061833
[Accessed January 30, 2018]

Name:
Takuya Fujinaga

Affiliation:
Department of Life Science and Systems Engineering, Kyushu Institute of Technology

Address:
2-4 Hibikino, Wakamatsu-ku, Fukuoka 808-0196, Japan

Brief Biographical History:

2016 Received Bachelor's degree, Department of Mechanical Information Science and Technology, Computer Science and Systems Engineering, Kyushu Institute of Technology
2018 Received Master's degree, Department of Human Intelligence Systems, Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology
2018- Doctoral Candidate, Department of Life Science and Systems Engineering, Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology

Membership in Academic Societies:

• The Japan Society of Mechanical Engineers (JSME)

Name:
Shinsuke Yasukawa

Affiliation:
The University of Tokyo
Recreation Lab, Inc.

Address:
4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan
4F Otemachi Bldg., 1-6-1 Otemachi, Chiyoda-ku, Tokyo 100-0004, Japan

Brief Biographical History:
2014-2017 Research Associate, Kyushu Institute of Technology
2017 Received Ph.D., Division of Electrical, Electronic and Information Engineering, Osaka University
2017- Researcher, Recreation Lab, Inc.
2017- Research Associate, The University of Tokyo

Main Works:
• "A Vision Sensor System with a Real-Time Multi-Scale Filtering Function," Int. J. of Mechatronics and Automation, Vol.4, No.4, pp. 248-258, 2014.
• "Real-Time Object Tracking Based on Scale-Invariant Features Employing Bio-Inspired Hardware," Neural Networks, Vol.81, pp. 29-38, 2016.

Membership in Academic Societies:
• The Robotics Society of Japan (RSJ)
• The Society of Instrument and Control Engineers (SICE)

Name:
Binghe Li

Affiliation:
Department of Brain Science and Engineering, Kyushu Institute of Technology

Address:
2-4 Hibikino, Wakamatsu-ku, Fukuoka 808-0196, Japan

Brief Biographical History:
2011 Received Bachelor's degree, Department of Systems Design and Informatics, Computer Science and Systems Engineering, Kyushu Institute of Technology
2013 Received Master's degree, Department of Brain Science and Engineering, Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology
2013- Doctoral Candidate, Department of Brain Science and Engineering, Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology

Membership in Academic Societies:
• The Robotics Society of Japan (RSJ)

Name:
Kazuo Ishii

Affiliation:
Professor, Department of Human Intelligence Systems, Kyushu Institute of Technology

Address:
2-4 Hibikino, Wakamatsu-ku, Kitakyushu, Fukuoka 808-0196, Japan

Brief Biographical History:
1996- Researcher, Institute of Industrial Science, The University of Tokyo
1996-1998 Assistant Professor, Kyushu Institute of Technology
1998-2011 Associate Professor, Kyushu Institute of Technology
2011- Professor, Kyushu Institute of Technology

Main Works:
• "Enhancement of deep-sea floor images obtained by an underwater vehicle and its evaluation by crab recognition," J. of Marine Science and Technology, Vol.22, Issue 4, pp. 758-770, 2017.

Membership in Academic Societies:
• The Institute of Electrical and Electronics Engineers (IEEE)
• The Japan Society of Mechanical Engineers (JSME)
• The Robotics Society of Japan (RSJ)
• The Institute of Electrical and Electronics Engineers (IEEE)