Integration of Color and Range Images for Plant Recognition

Takeshi Nishida, Shuichi Kurogi, Tomonari Kitamori, Ayako Eguchi and Yasuhiro Fuchikawa
Department of Control Engineering
Kyushu Institute of Technology, Kitakyushu 804-8550 Japan
Tel: +81-93-884-3193, Fax: +81-93-861-1159
E-mail: nishida@cntl.kyutech.ac.jp

Abstract: This paper presents an image processing system which integrates color and range images for plant recognition. The system consists of a pan-tilt-zoom color camera and a horizontal line-scanning laser rangefinder which can rotate vertically. The color and range images captured by these devices are integrated and fused so that the pixels of both images are related. At the same time, the color and the range (distance) data are used for extracting and classifying the parts of the plant, where a competitive neural network is used for classifying leaves and stalks. We show the experimental results where the present system could recognize the parts of the plant.

1. Introduction

In recent years, many practical and automatic plant recognition methods for weed removal system or automatic management of crops have been proposed [1]–[7]. The methods proposed in early researches [1]–[3] are based only on color image processing where a specific plant is detected by the color and crops and weeds are discriminated by 2-dimensional leaf shape analysis. However, since the main purpose of these studies is for spot spraying agricultural chemicals around target weeds in a limited area, the measurement of the individual position of the target plant is not taken into consideration. Furthermore, the methods are supposed to be used under limited situation, such as the case where the sprout time of target plants is controlled and the plants have comparatively few individual differences. On the other hand, a method to discriminate 30 or more varieties of wild grasses via image processing has been proposed [5], however, since this method requires the photo images, where a flower or a leaf of wild grass should be shapely extracted, it is difficult to apply the method to the living wild grass.

Over the past few years, new methods using a camera and a laser rangefinder (LRF) have been developed [6], [7], where the methods measure the accurate positions of crops and the degree of maturity of fruits by combining color and range images. However, since a fixed camera is used in these researches, the view of the camera is restricted.

Therefore, in this study, we have developed a system consists of a PTZ (pan-tilt-zoom) color camera and a horizontal line-scanning LRF which can rotate vertically by a suspension unit similar to those of [8], [9]. Furthermore, we propose a method using CNN (Competitive Neural Network) [10] for recognizing and discriminating the parts of a target plant. The overall flow of the present system is as follows (see Fig.1); first, the color image is captured and a pre-processing to extract the plant from background is carried out based on the hue value of the plant. At the same time, the 3-dimensional scanning by the LRF and a pre-processing to produce the range image are carried out. Next, these images are integrated and then the pixels of those images are related. Further, the range image is clustered into sub-areas according to the range data, where the range values of the pixels in each sub-area are close and the number of the pixels of each sub-area is larger than a certain constant. Finally, the parts of the target plant is discriminated into leaves and stalks through pattern matching by the CNN.

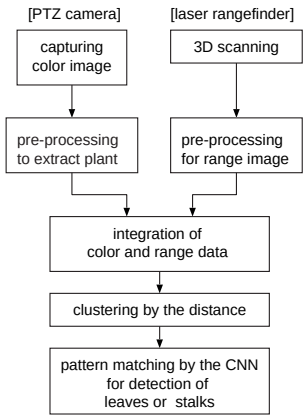


Figure 1. Schematic diagram of the present system.

2. Image Sensor System

We have constructed the sensor system shown in Fig.2 and the devices are placed as shown in Fig.3. Here, (X_L, Y_L, Z_L) and (X_C, Y_C, Z_C) are the LRF and the camera coordinates system, respectively.

2.1 PTZ Color Camera

(1) Capturing Color Image

A 24bit RGB color image $p(\boldsymbol{x})$ is captured by the PTZ color camera, where $p(\boldsymbol{x})$ represents the color or hue at

$$\boldsymbol{x} = \left\{ (i,\ j)^T \in \mathbb{R}^2 | i = 1, \cdots, N_x; j = 1, \cdots, N_y \right\}. \quad (1)$$

As a pre-processing for extracting the region of the target plant, we employ the modified hue transformation described bellow, and binarization, and noise elimination.

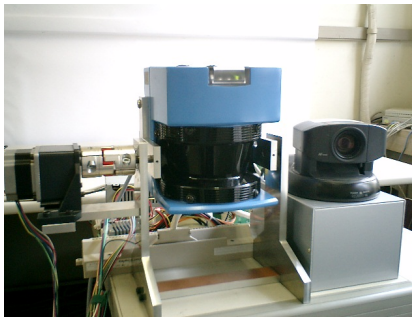


Figure 2. The sensor system consisting of the LRF (center) mounted on the suspension unit, the stepping motor (left), and the PTZ color camera (right).

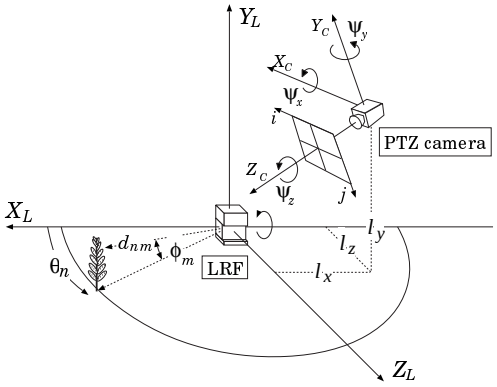


Figure 3. Relation between the coordinate systems of the LRF and the PTZ color camera.

(2) Modified Hue

Since the RGB values of the captured images tend to change by the strength of the surrounding light sources, we transform the color image into the hue image, i.e. each hue of the pixel (i, j) is calculated by

$$h_{ij} = \cos^{-1}\left(\frac{(r_{ij}-g_{ij})+(r_{ij}-b_{ij})}{2\left\{(r_{ij}-g_{ij})^2+(r_{ij}-b_{ij})(g_{ij}-b_{ij})\right\}^{1/2}}\right), \tag{2}$$

where (r_{ij}, g_{ij}, b_{ij}) are RGB values, and the hue ranges from 0[deg] to 360[deg]. However, according to [4], no value of hue was found within the range of 146[deg] to 228[deg] for any field surface, so that the modified hue has been proposed where hues grater than 180[deg] are represented as negative angles in order to eliminate the discontinuity around the red hue area (around 0[deg] and 360[deg]), and it has been shown that the best separation of plants and the background occurred with the modified hue among a number of color indices. We use the modified hue given by

$$h_{ij} := \begin{cases} h_{ij} - 180 & \text{if } h_{ij} > 180, \\ h_{ij} + 180 & \text{otherwise,} \end{cases} \tag{3}$$

which also has no discontinuity around the red hue area and has nonnegative value. An example of images is

(a) (b)

Figure 4. Example of a camera image of a plant. (a) A color image (240×320[pixel]). The rectangle region is a target plant region calculated by (b) a preprocessed image based on modified hue value.

shown in Fig.4(a), where the rectangular region involving the target plant, which we call the target plant region, is obtained via the modified hue method. And then, we obtain Fig.4(b) by means of binarizing the image with respect to the modified hue values of the target plant, and noise reduction via moving average filter and quantization filter. The target plant region is used for the integration with the following range data.

2.2 LRF

We use the LRF (SICK LMS 200) which basically scans the horizontal two-dimensional plane to measure the distance to the object, where the maximum measurable distance is 8191[mm], and the maximum scanning range is 180[deg] with angular resolution 0.25[deg]. In order to scan three-dimensional space, we have designed to make a suspension unit for rotating the LRF vertically by means of a geared stepping motor with angle resolution 0.05[deg]. Thus, we can obtain the three-dimensional position $\boldsymbol{x}^L_{nm}$ from the measured distance d_{nm} for the horizontal angle θ_n and the vertical rotation angle ϕ_m, as follows:

$$\boldsymbol{x}^L_{nm} = \begin{pmatrix} x^L_{nm} \\ y^L_{nm} \\ z^L_{nm} \end{pmatrix} = \begin{pmatrix} d_{nm}\cos\phi_m\cos\theta_n \\ d_{nm}\sin\phi_m \\ d_{nm}\cos\phi_m\sin\theta_n \end{pmatrix}, \tag{4}$$

where we use the angles given by

$$\theta_n = 0.25n + 40[\text{deg}] \quad (n = 0, 1, \cdots, 400), \tag{5}$$

$$\phi_m = 0.25m - 50[\text{deg}] \quad (m = 0, 1, \cdots, 400). \tag{6}$$

The example of the range image of the same target plant as given in Fig.4 is shown in Fig.5, where the distance from 200 to 1200[mm] is digitized into 256 gray levels.

3. Integration and Clustering of Measurement Data

3.1 Integration of color and range image

We show two methods of integrating color and range images; one is to match the pixels of both images directly, and the other is to map the target plant region of the color image to the range image.

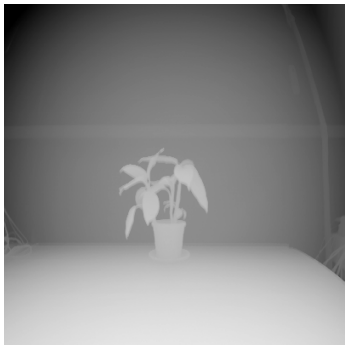
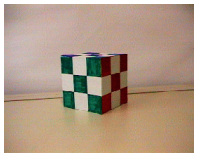
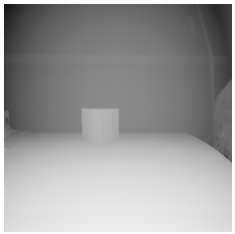


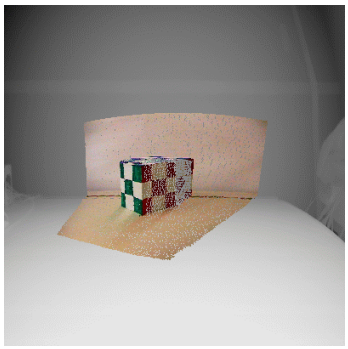
Figure 5. Range image of a plant(401 × 401 [pixel]).



(a) Color image.



(b) Range image.



(c) Integrated image.

Figure 6. Example of the pixel matching integration of images.

(1) Pixel Matching

The position $\boldsymbol{x}_{nm}^C = (x_{nm}^C, y_{nm}^C, z_{nm}^C)^T$ on the camera coordinate system can be represented by the position $\boldsymbol{x}_{nm}^L$ on the LRF coordinate system as follows,

$$\boldsymbol{x}_{nm}^C = \boldsymbol{R}(\boldsymbol{\psi}) \cdot \boldsymbol{x}_{nm}^L + \boldsymbol{l}, \tag{7}$$

where $\boldsymbol{R}(\boldsymbol{\psi}) \in \mathbb{R}^{3\times 3}$ is three-dimensional rotation matrix with respect to $\boldsymbol{\psi} = (\psi_x, \psi_y, \psi_z)^T$, where ψ_x, ψ_y and ψ_z are the rotation angles around the X_C, Y_C and Z_C axes (see Fig.2), and $\boldsymbol{l} = (l_x, l_y, l_z)^T$ is the position of the camera from the LRF. Through the perspective projection, the position (i, j) on the camera image is represented as follows,

$$(i,\ j) = \left(f_c \frac{x_{nm}^C}{z_{nm}^C} + \frac{N_x}{2},\ -\left(f_c \frac{y_{nm}^C}{z_{nm}^C} + \frac{N_y}{2} \right) \right), \tag{8}$$

where f_c is the focal length. Thus, the color of the position (θ_n, ϕ_m) on the range image is supposed to be the same color of the corresponding position (i, j) on the color image, where (i, j) corresponds to (θ_n, ϕ_m) through Eqs.(4)–(8). Here, note that when more than one position on the range image correspond to a position (i, j) on the color image, or some positions corresponding to range pixels are occluded in the camera image, we set the color of (i, j) to the pixel (θ_n, ϕ_m) with the minimum d_{nm}.

An example of the above integration of two images for a color cube is shown in Fig.6, where the image (c) originally is the range image (b) but the pixels corresponding to the color image (a) is colored. Although most of range pixels look like colored correctly, some errors have occurred, and we here focus on two types of errors. One is on the red face of the cube, where the pixels of the range image cannot match one-to-one to the pixels of the color image. To solve this problem, it seems to be a good idea to increase the resolution of the color image, which, however, is not appropriate for the other type of errors. Namely, the second type of errors has occurred at the pixels which are considered to be occluded in the color image although we have employed the occlusion process as described above. This problem may be removed when the resolution of the range image is higher than that of the color image, which however is not appropriate for the first problem. There may be several solutions of these problems, we in this report present the following integration methods for extracting the parts of the plants.

(2) Region Extracting

This approach is to map the target plant region (see Sect.2.1) in the color image to the range image, which is processed by the next step shown in the following section.

Let us consider Fig.7(a), where the range data of the target exists inside the pyramid with the vertex on the origin of the camera coordinate system corresponding to the angle of view of the target plant region.

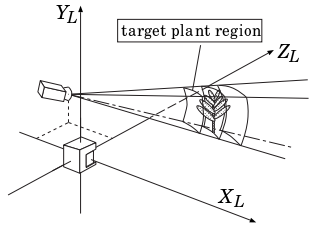
From the relation shown in Fig.7(b), we derive the equation of the correspondence between the angle of view of the region and the direction of the scanning lay of the LRF on the X_L-Z_L plane. The lines $Z_L^{(1)}$, $Z_L^{(2)}$ and $Z_L^{(3)}$ in Fig.7(b) are represented as follows,

$$Z_L^{(1)} = (X_L - l_x)\tan\alpha + l_z, \tag{9}$$
$$Z_L^{(2)} = (X_L - l_x)\tan\beta + l_z, \tag{10}$$
$$Z_L^{(3)} = X_L \tan\theta_n, \tag{11}$$

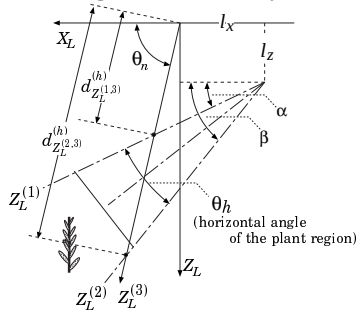
where

$$\alpha = \psi_y - \theta_c - \theta_h/2, \tag{12}$$
$$\beta = \psi_y - \theta_c + \theta_h/2, \tag{13}$$

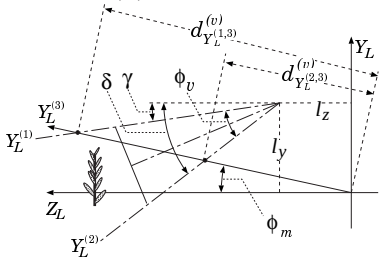
and θ_c is the orientation of the center of the target plant region, and θ_h is the horizontal range (angle) of the target plant region, which are obtained as follows (see



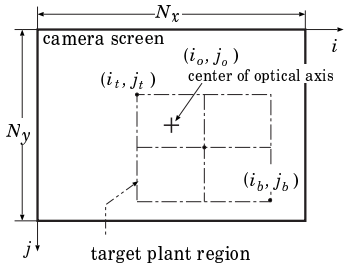
(a) Relation between the camera and the range coordinate system.



(b) Top view.



(c) Side view.



(d) The plant region on the camera screen.

Figure 7. Relation between the coordinate system of the LRF, the camera, and the camera screen.

Fig.7(d)),

$$\theta_c = \left(\frac{i_b + i_t}{2} - i_o\right)\frac{\Theta}{N_x}, \quad (14)$$

$$\theta_h = (i_b - i_t)\frac{\Theta}{N_x}, \quad (15)$$

where $(i_t,\ j_t)$ and $(i_b,\ j_b)$, respectively, are the positions of the upper right and the lower left pixels of the target plant region, and $\Theta = 0.977$[rad] $\simeq 56$[deg] is the horizontal view angle of the camera. From the above relations, the distances from the LRF to the crossing of $Z_L^{(1)}$ and $Z_L^{(3)}$, and that of $Z_L^{(2)}$ and $Z_L^{(3)}$ are expressed as follows,

$$d^{(h)}_{Z_L^{(1,3)}} = \frac{l_x \tan\alpha - l_z}{\cos\theta_n\,(\tan\alpha - \tan\theta_n)}, \quad (16)$$

$$d^{(h)}_{Z_L^{(2,3)}} = \frac{l_x \tan\beta - l_z}{\cos\theta_n\,(\tan\beta - \tan\theta_n)}, \quad (17)$$

where we set

$$\begin{cases} d^{(h)}_{Z_L^{(1,3)}} := \infty & \text{if } d^{(h)}_{Z_L^{(1,3)}} \le l_z \\ d^{(h)}_{Z_L^{(2,3)}} := \infty & \text{if } d^{(h)}_{Z_L^{(2,3)}} \le l_z \end{cases} \quad (18)$$

In the same way, the lines in Fig.7(c) are represented as follows,

$$Y_L^{(1)} = (Z_L - l_z)\tan\gamma + l_y, \quad (19)$$

$$Y_L^{(2)} = (Z_L - l_z)\tan\delta + l_y, \quad (20)$$

$$Y_L^{(3)} = Z_L \tan\phi_m, \quad (21)$$

where,

$$\gamma = \psi_x - \phi_c + \phi_v/2, \quad (22)$$

$$\delta = \psi_x - \phi_c - \phi_v/2, \quad (23)$$

$$\phi_c = \left(\frac{j_b + j_t}{2} - j_o\right)\frac{\Phi}{N_y}, \quad (24)$$

$$\phi_v = (j_b - j_t)\frac{\Phi}{N_y}, \quad (25)$$

and $\Phi = 0.733 \simeq 42$[rad] is the vertical view angle of the camera. Moreover the distances of the crossing of $Y_L^{(1)}$ and $Y_L^{(3)}$, and that of $Y_L^{(2)}$ and $Y_L^{(3)}$ on the Y_L-Z_L vertical plane shown in Fig.7(c) are derived as follows,

$$d^{(v)}_{Y_L^{(1,3)}} = \frac{l_z \tan\gamma - l_y}{\cos\phi_m\,(\tan\gamma - \tan\phi_m)}, \quad (26)$$

$$d^{(v)}_{Y_L^{(2,3)}} = \frac{l_z \tan\delta - l_y}{\cos\phi_m\,(\tan\delta - \tan\phi_m)}, \quad (27)$$

where we set

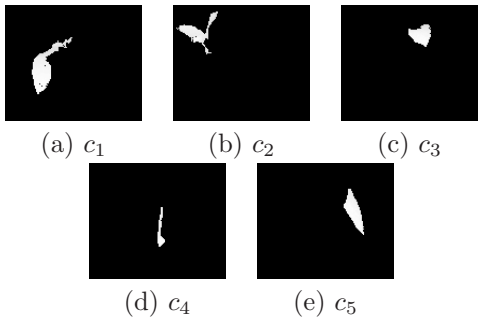
$$\begin{cases} d^{(v)}_{Y_L^{(1,3)}} := \infty & \text{if } d^{(v)}_{Y_L^{(1,3)}} \le l_z, \\ d^{(v)}_{Y_L^{(2,3)}} := \infty & \text{if } d^{(v)}_{Y_L^{(2,3)}} \le l_z. \end{cases} \quad (28)$$

If the range data $(\theta_n, \phi_m, d_{nm})$ fulfill the following equations,

$$\begin{aligned} d^{(h)}_{Z_L^{(1,3)}} < d_{nm} < d^{(h)}_{Z_L^{(2,3)}}, \\ d^{(v)}_{Y_L^{(1,3)}} < d_{nm} < d^{(v)}_{Y_L^{(2,3)}}, \end{aligned} \quad (29)$$

the pixel (θ_n, ϕ_m) of the range image is supposed to be in the target plant region. By applying the above integration method to the target plant region shown in Fig.4(a) and the range image shown in Fig.5, we have got the result as shown in Fig.8. The integration result is useful for extracting the features of the target plant as shown the next section.

Figure 8. Example of a extracted range image corresponding to the target plant region of the camera image (130 × 113 [pixel]).



(a) c_1 (b) c_2 (c) c_3 (d) c_4 (e) c_5

Figure 9. Clustered range images.

3.2 Clustering of range image

In order to recognize a feature of a target plant, we extract several sub-areas from the range image integrated via the method presented in the previous section, where each sub-area consists of pixels adjacent each other and $|d_{nm} - d_{n'm'}|$ of the adjacent pixels (θ_n, ϕ_m) and $(\theta_{n'}, \phi_{m'})$ is less than a certain value (6[mm]), and the number of the pixels in the sub-area are more than a certain number (100[pixel]), where the wall and the bowl were deleted on the basis of the number of data points. The extracted sub-areas are shown in Fig.9.

4. Competitive Neural Network

For detecting the features of the target plant whose shape and size change with view points, we employ the CNN [10] (see Fig.10) which can achieve the invariant recognition to linear and/or nonlinear coordinate transformations.

4.1 Structure

The CNN consists of a number of clusters, each of which has competitive cells with a N^2-dimensional

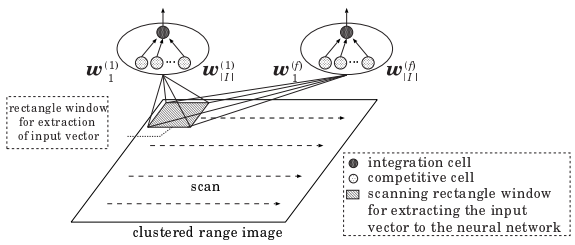
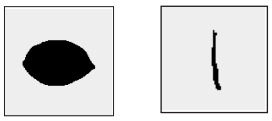


Figure 10. Competitive neural network.



(a) $q^{(1)}(\boldsymbol{x})$ (b) $q^{(2)}(\boldsymbol{x})$

Figure 11. Template patterns of (a) a leaf and (b) a stalk are binary image consists of 64 × 64 pixels. These template patterns were made from a real leaf and stalk of the target plant.

weight vector $\boldsymbol{w}_k^{(f)} \in \mathbb{R}^{N^2}$ $(k \in K,\ f \in F,\ N = 64)$ and an integration cell, where $K = \{1, \cdots, |K| = 200\}$ and $F = \{1, \cdots, |F| = 2\}$ are the sets of indices of the competitive cells and the clusters (or the features), respectively. The input to the CNN at the learning phase is the transformed template patterns $q^{(f)}(\boldsymbol{x})$ which are shown in Fig.11, and that at the recognition phase is the image of the extracted sub-area.

4.2 Learning

The weight vectors $\boldsymbol{w}_k^{(f)} = \left(w_{k1}^{(f)}, \cdots, w_{kN^2}^{(f)}\right)^T$ $(k \in K)$ of the competitive cells in the fth cluster which corresponds to the fth feature are trained with the transformed and normalized patterns in

$$P_G^{(f)} = \left\{ \boldsymbol{q}^{(f)}(\boldsymbol{g}(\boldsymbol{x})) \Big/ \|\boldsymbol{q}^{(f)}(\boldsymbol{g}(\boldsymbol{x}))\| \Big| \boldsymbol{g} \in G \right\},$$

by means of the competitive algorithm described below. Here, the projective transformation $\boldsymbol{g}(\cdot)$ is nonlinear coordinate transformation given by

$$\boldsymbol{g}(\boldsymbol{x}) = \frac{A\boldsymbol{x} + \boldsymbol{b}}{\boldsymbol{p}^T\boldsymbol{x} + 1} \tag{30}$$

where

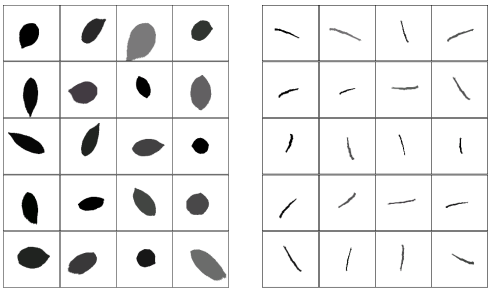
$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},\ \boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix},\ \boldsymbol{p} = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix},$$

$a_{11}, a_{12}, a_{21}, a_{22}, b_1, b_2, p_1, p_2 \in \mathbb{R}$. In order to obtain the invariant recognition ability for various coordinate transformations, the CNN employs the CRL (Competitive Reinitialization Learning) method for vector quantization (see [11] for details of CRL).

The weight vectors after the training in the 1st and 2nd cluster are shown in Fig.12, which shows that the network have learned various projective coordinate transformation patterns, where the weight vectors in each cluster, respectively, have been trained with 1000 randomly generated projective transformation patterns of the template patterns.

4.3 Recognition

At the recognition phase, the input vector $\boldsymbol{x}_s \in \mathbb{R}^{N^2}$ $(s = 1, 2, \cdots)$ to the network is extracted with a $N \times N$ [pixel] rectangle window which moves a few pixel at a time from the clustered range image c_t $(t =$



(a) $\boldsymbol{w}_1^{(1)} - \boldsymbol{w}_{20}^{(1)}$ (b) $\boldsymbol{w}_1^{(2)} - \boldsymbol{w}_{20}^{(2)}$

Figure 12. Weight vectors trained with the CRL. (a) leaves and (b) stalks.

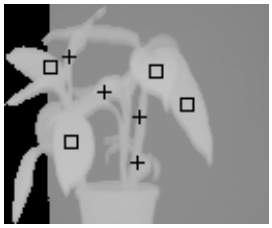


Figure 13. Recognition result. The marks of rectangle and plus represent the positions of leaves and stalks, respectively.

$1, 2, \cdots$). When the $\boldsymbol{x}_s$ is entered to the net, each cell calculates the inner product between the weight vector and the input vector, and the integrated cell selects and outputs the value of the maximum inner product as follows,

$$y^{(f)}(c_t) = \max_{k,s} \left(\frac{\boldsymbol{x}_s}{\|\boldsymbol{x}_s\|} \cdot \boldsymbol{w}_k^{(f)} \right). \tag{31}$$

Thus, this equation shows that the inner product gives the maximum value when $\boldsymbol{x}_s$ is contains $q^{(f)}(\boldsymbol{g}(\boldsymbol{x}))$, and the number $\bar{f}$ with the maximum output $y^{(\bar{f})}(c_t)$ is judged to be the number of the recognized feature. Here, we decide whether or not the input pattern $\boldsymbol{x}_s$ matches with $q^{(f)}(\boldsymbol{g}(\boldsymbol{x}))$ by means of the condition:

$$y^{(f)}(c_t) \geq s_{th}, \tag{32}$$

where s_{th} represents a threshold of the matching ratio, and we set $s_{th} = 0.8$ in the following experiment.

As a result of applying the above recognition processing by the network to the clustered range images, c_3 and c_5 have been recognized as a leaf, c_4 has been recognized as a stalk, and c_1 and c_2 have been recognized a leaf and a stalk, respectively. The figure which unified these results is shown in Fig.13. This result shows that the features of the target plant have been recognized correctly. Moreover, since this recognition processing was performed on the range data image, the distance to each recognized feature part can also be known.

5. Concluding Remarks

We have presented an image processing system which consists of a LRF and a PTZ camera for capturing a color and a range images of a target plant, and estimated two kinds of the integration method of those measurement data. Moreover, we have presented the recognition method to classify leaves and stalks of a target plant by using a competitive neural network.

References

[1] M. Chapron, L. Martin-Chefson, L. Assemat and P. Boissard, "A multiresolution weed recognition method based on multispectral image processing," *Proc. of SCIA '99*, pp.529–534, 1999.

[2] L. Tang, G. E. Meyer, K. Von Bargen, D. A. Mortense, "Color indices for weed identification under various soil, residue, and lighting conditions," *Trans. of the ASAE*, Vol.38, no.1, pp.259–269, 1995.

[3] G. E. Meyer, T. Mehta, M. F. Kocher, D. A. Mortense, A. Samal, "Textural imaging and discriminant analysis for distinguishing weeds for spot spraying," *Trans. of the ASAE*, Vol.41, no.4, pp.1189–1197, 1998.

[4] D. M. Woebbecke, G. E. Meyer, K. Von Bargen, D.A.Mortense, "Color indices for weed identification under various soil, residue, and lighting conditions," *Trans. of the ASAE*, Vol.38, no.1, pp.259–269, 1995.

[5] T. Saitoh and T. Kaneko, "Automatic Recognition of Wild Flowers," *IEICE(D-II)*, Vol.J84-D-II, no.7, pp.1419–1429, 2001.

[6] K. Hatou, T. Sugiyama and Y. Hahimoto, "System Design for Plant Factory Operation Based on Three-dimensional Data Base (4) Recognition of 3-dimentional Image by Using Neural Network," *Journal of Shita*, Vol.8, no. 1, pp.49–53, 1996.

[7] H. Matsuura, K. Hatou, J. Yamashita and Y. Hashimoto, "Recognition by Robot Visual Sensor Based on Sensor Fusion -Fuse Range Image and 2-dimentional Image -," *Journal of Shita*, Vol.9, no. 2, pp.132–138, 1997.

[8] H. Surmann, K. Lingemann, A. Nüchter and J. Hertzberg," A 3D laser range finder for autonomous mobile robots, " Proc. of the 32nd ISR, pp.153–158, 2001.

[9] H. Surmann, K. Lingemann, A. Nüchter and J. Hertzberg, "Fast acquiring and analysis of three dimensional laser range data," *Proc. of VMV 2001*, pp.59–66, 2001.

[10] T. Nishida and S. Kurogi, "Multi-layered competitive net for pattern recognition invariant to coordinate transformations," *Journal of Japanese Neural Network Society(in Japanese)*, Vol.7, no4,106–114, 2000.

[11] T.Nishida, S.Kurogi and T.Saeki, "Adaptive vector Quantization using re-initialization method," *IEICE(D-II)*, Vol.J84-D-II, no.7, pp.1503–1511, 2001.