Counting of Pear Flower Buds in Images by Judging Acquisition Conditions and Matching Keypoints

Takumi Hanakawa*, Shintaro Nakatani*, Jaehwan Lee†, Eiji Morimoto†, Masashi Nishiyama* and Yoshio Iwai*,
*Graduate School of Sustainability Science, Tottori University, Japan
†Graduate School of Agricultural Science, Kobe University, Japan

***Abstract*—We propose a method for accurately counting pear flower buds from worm's-eye view images acquired from a mobile ground-based camera system, using acquisition condition judgment and keypoint matching. To develop this method, we considered the following two issues. First, the appearance of flower buds acquired at each time point varies with the illumination and camera parameters. In this case, using a single detector of flower buds frequently reduces the count accuracy. Second, the extent of overlap in the camera field of view among acquisition time points varies with the camera system setting. In this case, the counting of identical flower buds is duplicated if only the detection process is performed at each time point. To accurately count flower buds, the proposed method detects the candidate flower buds in the worm's-eye view images acquired at each time point using a suitable detector selected by the acquisition condition judgment and identifies identical flower buds using keypoint matching in image pairs between time points. Experimental results demonstrated that our method counts flower buds more accurately than the comparative methods that use a single detector.**

***Index Terms*—Worm's-eye view images, Pear flower buds, Counting, Acquisition condition, Keypoint matching**

I. Introduction

With the introduction of smart agriculture in pear orchards, the acquisition of information on the growth of pear trees must be considered. Specific information on the growth of pear trees includes the number of fruit, the weight of fruit, the number of flower buds, and the direction of branch growth. Among these variables, the number of fruit is vitally important information for pear growers, and the number of flower buds provides crucial information on growth leading up to fruit development, as described previously [1]. To manage fruit production in a pear orchard, it is useful to determine the number of flower buds in early spring and to compare this with the number of fruit in fall. Furthermore, accumulation of a large amount of such information may enable future prediction of the number of fruit. In previous studies to obtain information on the growth of pear trees, Parico et al. [2] and Baerdemaeker et al. [3] proposed methods to count pear fruit. However, these methods [2], [3] consider only pear fruit and do not assess pear flower buds. Recently, Deckers et al. [4] proposed a method to count pear flower buds using a multispectral camera system. In this paper, we describe the use of a simple and inexpensive color camera to acquire images of pear flower buds.

The present study evaluated a method for accurately counting flower buds from time-series worm's-eye view images acquired by a mobile ground-based color camera system. A simple approach is to apply an object detector [5]–[7] with deep learning techniques to the worm's-eye view image at each time point to detect candidate regions of pear flower buds and to count the number of candidate regions. For example, Farjon et al. [8] applied Faster R-CNN [7] to detect apple flower buds. However, simple application of such detectors reduces the accuracy of the counts because of variation generated when acquiring images of pear flower buds.

We considered the following two crucial issues that arise from variation in the image acquisition conditions. First, variation attributable to factors such as illumination and camera parameters reduces the accuracy of flower bud counts at each time point. Specifically, the apparent color of flower buds in images varies with the weather-dependent illumination condition. In addition, the apparent size of flower buds in the images varies depending on the lens and resolution parameters. Second, variation resulting from factors such as the camera system setting cause overlapping detection of identical flower buds between time points, which reduces the accuracy of flower bud counts. Specifically, even if the camera frame rate is constant, the interval between the mobile camera-system positions varies with the ground surface conditions, causing overlap in the view areas where the counting of identical flower buds may be erroneously duplicated between successive time points.

In this paper, we propose a method to accurately count flower buds in worm's-eye view images by selecting a detector suitable for the appearance of flower buds at each time point using acquisition condition judgment and determining identical flower buds using keypoint matching in image pairs between successive time points of image acquisition. To address the first issue described in this section, our method introduces acquisition condition judgment, to cope with variation in factors such as illumination and camera parameters, to improve the accuracy of counts at each time point. With regard to the second issue mentioned in this section, our method introduces keypoint matching to cope with variation, in factors such as the camera system settings, to improve the accuracy of counts between successive time points. Experimental results demonstrated that the proposed method of counting pear flower buds is more accurate than the comparative methods, which rely on a single flower-bud detector.

The proposed method used a mobile ground-based color

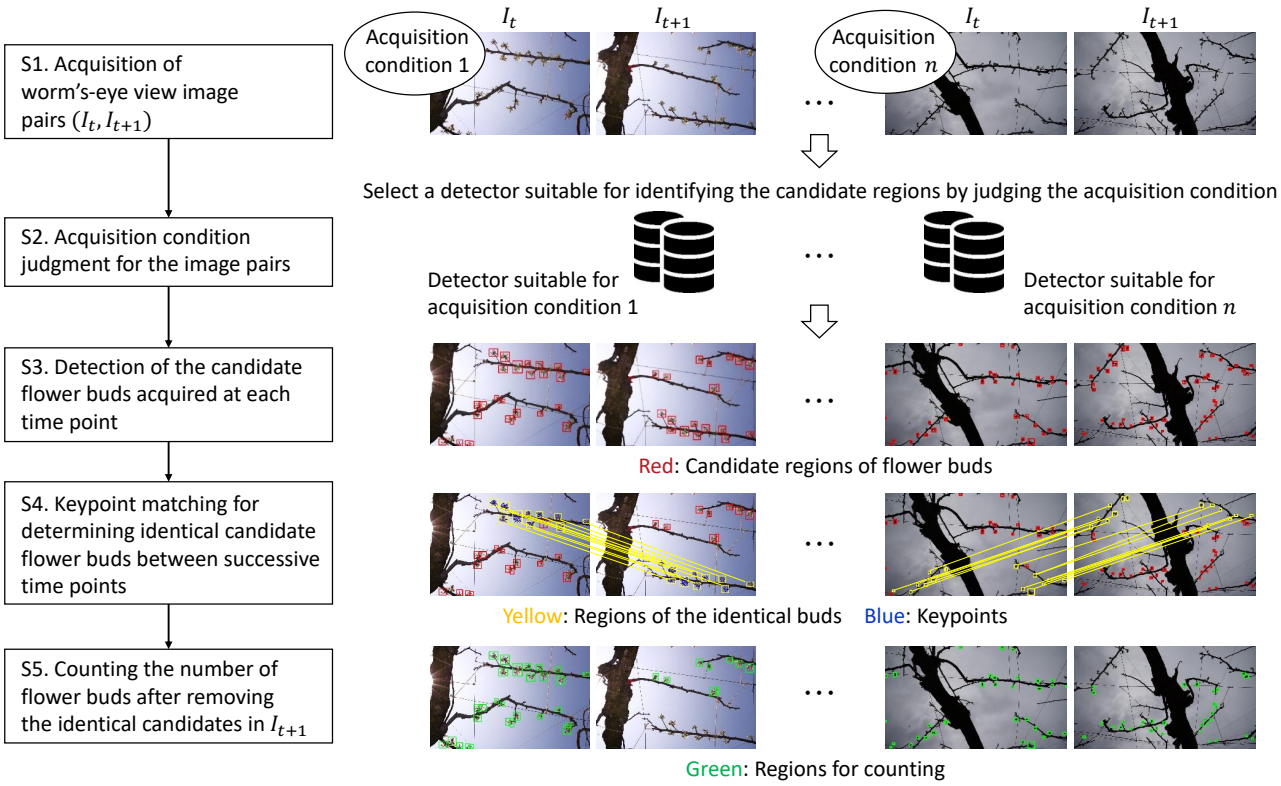

Fig. 1. Overview of our method.

camera system to acquire worm's-eye view images. Generally, aerial photography by unmanned aerial vehicles is used in smart agriculture. However, the pear orchard selected for the present study is located near the sea and on a small mountain slope. Therefore, coastal wind from the ocean strongly affects aerial image acquisition, making it impossible for an unmanned aerial vehicle to acquire images stably. We chose to use a ground-based camera system owing to its adaptability and effectiveness in the conditions specific to the pear orchard.

II. Our Method

We provide an overview of our method in Section II-A and describe the main components of our method in Sections II-B through II-D.

A. Overview

Figure 1 presents an overview of the proposed method for counting pear flower buds using worm's-eye view images. First, in step S1, the image pair (I_t, I_{t+1}) is acquired using a mobile ground-based color camera system. An image pair consists of a worm's-eye view image I_t acquired at time point t and a worm's-eye view image I_{t+1} acquired at time point $t+1$.

Next, in step S2, a detector for detection of candidate flower-bud regions is determined. To address the first issue described in Section I, we use image classification to determine the acquisition condition and select a detector suitable for the appearance of the flower buds in the images I_t and I_{t+1}. Step S2 aims to select a detector that can achieve high count accuracy, even if the acquisition condition of the worm's-eye view images varies as a result of illumination and camera parameters.

Next, in step S3, the candidate flower-bud regions are detected using the detector selected in step S2 for the image pair (I_t, I_{t+1}). However, the use of step S3 alone may cause duplicate counting owing to the second issue described in Section I.

Next, in step S4, the identical candidate flower-bud regions in image pair (I_t, I_{t+1}) are determined using keypoint matching. Specifically, an identical flower bud is determined when a keypoint in a candidate region of I_t corresponds to a keypoint in a candidate region of I_{t+1}. This step solves the second issue caused when the view area overlaps in the image pair (I_t, I_{t+1}). Step S4 aims to prevent duplicate counting of flower buds, even if the acquisition condition of the worm's-eye view images varies as a result of the camera system setting.

Finally, in step S5, identical flower buds are excluded from I_{t+1} using the corresponding candidate regions and the total number of remaining candidate regions is counted in the image pair (I_t, I_{t+1}) as the number of flower buds.

B. Acquisition condition judgment for the image pair

When the image acquisition condition changes, the appearance of the flower buds in images changes owing to factors such as illumination and camera parameters, as described in Section I. In this case, using only a single detector leads to the first issue, which is that the accuracy of pear flower-bud

counting decreases. In step S2, the acquisition conditions of the input worm's-eye view images I_t and I_{t+1} are determined and a detector suitable for the appearance of flower buds in those images is selected. The method uses an existing image classification technique, ResNet [9], to determine the acquisition conditions. Visually similar worm's-eye view images are collected for each acquisition condition and are used to train the ResNet model. The method uses the ResNet50 model as the deep neural network architecture, resamples the size of the worm's-eye view image to 224×224 pixels, and trains the model with fine-tuning.

C. Detection of the candidate flower buds in images acquired at each time point

In step S3, the method detects candidate flower-bud regions from the worm's-eye view image using the detector selected in the acquisition condition judgment at step S2. High accuracy can be expected by using deep learning techniques [5], [6] as a detector. Even with these techniques, however, one problem requires attention. If the size of the candidate flower-bud region is relatively small compared with the size of the input image, the detection accuracy decreases, even if the existing deep learning techniques are used. Among the detectors used in this paper, we here consider SSD [5]. In the SSD300 model, one example of SSD implementation, each image in the pair (I_t, I_{t+1}) is resized to 300×300 pixels before it is input to the deep neural network detector. In this case, the size of the median bounding box of the candidate flower-bud region in the images is 9×13 pixels. Thus, the size of the resized bounding box is tiny, making it difficult to detect candidate regions.

To improve the accuracy of detecting candidate flower buds of small size, the method divides each worm's-eye view image of the pair I_t and I_{t+1} into smaller parcels and inputs them to the SSD model. Let I_t^b and I_{t+1}^b $(b = 1, \ldots, B)$ be images divided into B parcels. This procedure makes the bounding box size of I_t^b and I_{t+1}^b B times larger than the bounding box size of flower buds of I_t and I_{t+1}. The method alleviates the difficulty of detecting small candidate flower-bud regions by dividing the worm's-eye view images of image pair (I_t, I_{t+1}) into parcels.

D. Keypoint matching for determining identical candidate flower buds between successive time points

When the acquisition condition changes as a result of the camera system setting, overlap in the camera field of view occurs between time points as described in Section I. This issue results in an overlapping view area in the field of view between the images I_t and I_{t+1}. In this case, the detection process at each time point alone causes the second issue: identical flower buds that appear in duplicate in the overlapping view area at time points t and $t+1$ will be counted redundantly. Step S4 of the method uses keypoint matching to determine identical flower buds in image pair (I_t, I_{t+1}). Keypoint matching is a technique for extracting keypoints from an image pair and corresponding the keypoints to those with similar local descriptors between images. In this paper, we apply the SuperPoint [10] and SuperGlue [11] networks, which are used by Hanakawa et al. [12] for identification of flower buds. However, the present paper adds a novel improvement to accurately perform keypoint matching. SuperPoint finds keypoints and local descriptors from images using a feature map inside a neural network feature extractor. A local descriptor is a feature vector representing a local region centered on a keypoint. After finding keypoints in each image of the image pair (I_t, I_{t+1}), SuperGlue corresponds the keypoints of I_t to those of I_{t+1}. SuperGlue uses a neural network model trained from various image pairs to retain corresponding keypoints between images and remove those that do not correspond. The proposed method uses the SuperPoint and SuperGlue network models with publicly available default parameter weights.

In step S4, we add a novel improvement in applying keypoint matching between time points. The method applies post-processing to the correspondence results to maintain the consistency of the positional relationships among identical flower-bud candidate regions. In pear orchards, the branches of the pear trees are adjusted so that the flower buds are at approximately the same height from the ground to facilitate fruit harvesting. Thus, the positional relationships between the candidate regions of identical flower buds remain approximately the same. The positional relationship here refers to the angle that represents the inclination of the straight line connecting the identical keypoints in the image pair (I_t, I_{t+1}) when the I_t is concatenated with I_{t+1} horizontally. The method calculates the angle θ between the keypoint coordinates $(x_t,\ y_t)$ of I_t and the keypoint coordinates $(x_{t+1},\ y_{t+1})$ of I_{t+1} using the following equation:

$$\begin{cases} \theta = \arctan(\frac{x_{t+1}+w-x_t}{|y_{t+1}-y_t|}) & (x_{t+1} \neq w - x_t, y_{t+1} \neq y_t) \\ \theta = \frac{\pi}{2} & (x_{t+1} = w - x_t, y_{t+1} \neq y_t) \\ \theta = 0 & (y_{t+1} = y_t) \end{cases} \tag{1}$$

where w is the width of each worm's-eye view image of image pair (I_t, I_{t+1}), and $\arctan(\ )$ is the arctangent function. For each image pair, the distribution of θ is calculated by finding the angles among all corresponding candidate regions. We define the range $[\overline{\theta} - \theta_m, \overline{\theta} + \theta_m]$ using the margin θ_m and the median $\overline{\theta}$ of the distribution of θ. The method counts flower buds by excluding the identical candidate regions that are outside of the range.

III. Experiments

We describe the datasets for the evaluation in Section III-A, the evaluation index in Section III-B, and the results of counting flower buds in Section III-C. We also present the visualization of each step of our method in Section III-D.

A. Datasets

To evaluate the accuracy of the method in counting flower buds, we acquired a worm's-eye view image dataset of pear trees. In this paper, we used a mobile ground-based camera system to acquire worm's-eye view images of pear trees.

(a) Camera system (b) Path of the camera system (indicated by red arrows)

Fig. 2. Camera setting.

Figures 2(a) and (b) show the set-up of the camera system and the running scenery, respectively. We acquired images of pear trees growing in the Hashimoto Garden in Fukube-cho, Tottori City, Tottori Prefecture, Japan. The details of the dataset for each acquisition condition are shown below.

- Dataset 1 (Acquisition condition 1)
  We acquired the images on April 3, 2020, from 10:28 to 16:39. Figure 3(a) shows examples of worm's-eye view images acquired in this condition. The weather was clear and almost cloudless. The image size was 6000×4000 pixels and the number of images acquired was 670. The total number of manually annotated flower-bud bounding boxes was 10832. The median width and height of the bounding boxes were 181 and 180 pixels, respectively.
- Dataset 2 (Acquisition condition 2)
  We acquired the images on March 25, 2021, from 09:30 to 13:40. Figure 3(b) shows examples of worm's-eye view images acquired in this condition. The weather was cloudy with continuous thick clouds. The image size was 1920×1080 pixels and the number of images acquired was 670. The total number of manually annotated flower-bud bounding boxes was 24533. The median width and height of the bounding boxes were 23 and 24 pixels, respectively.

Comparing the appearances of the worm's-eye view images between Fig. 3(a) and Fig. 3(b), the entire images are colorful in acquisition condition 1, whereas the images are almost black and white in acquisition condition 2. This is due to the difference in weather conditions, one of the acquisition conditions associated with illumination. The size of the flower buds in the image is larger in acquisition condition 1 than in acquisition condition 2. This is due to the difference in the lens parameters, one of the acquisition conditions associated with the camera parameters. Figure 4 shows enlarged images of the flower buds for each dataset. Thus, it can be seen that the colors affected by illumination and the degree of blurring influenced by resolution differ among the acquisition conditions.

B. Evaluation index of the accuracy of counting flower buds

The F-measure is generally used as an evaluation index for object detection. The F-measure can be applied to evaluate the accuracy of detecting candidate flower-bud regions at each time point. However, it does not consider overlapping counts between time points. Therefore, the F-measure cannot be directly applied as an evaluation index for the accuracy of counting flower buds in an image pair. In this paper, we consider improvement of the F-measure to evaluate the counting accuracy in image pairs. We here term it the Improved F-measure and explain its calculation in the following text.

When calculating a conventional F-measure, a bounding box is assigned as an annotated label for each target object in a given image. In contrast, when calculating the Improved F-measure in this paper, in each worm's-eye view image of a given image pair (I_t, I_{t+1}), a bounding box is assigned as an annotated label for each flower bud, and an annotated corresponding label is also assigned to the identical flower bud in the image pair. The Improved F-measure retains the annotated bounding boxes of I_t and excludes the annotated bounding boxes of I_{t+1} that correspond to those of I_t; thus, the annotated bounding boxes of I_{t+1} that overlap with those of I_t are removed.

To calculate the Improved F-measure, we calculate the True Positive (TP), False Negative (FN), and False Positive (FP), as well as the conventional F-measure using the annotated bounding boxes with duplicates removed within image pairs. Although the conventional F-measure is calculated for a single image, the Improved F-measure is simultaneously calculated for both images of a given image pair. The TP of the Improved F-measure is the number of annotated bounding boxes for which the predicted bounding boxes are correctly detected. In counting TP, we used the condition that the Intersection over Union between the annotated bounding box and the predicted bounding box was greater than or equal to 0.5. The FN of the Improved F-measure is the number of annotated bounding boxes for which no predicted bounding box was detected by mistake. The FP is the number of predicted bounding boxes obtained by mistake when there was no annotated bounding boxes. We calculated the Improved F-measure from the harmonic mean of Precision and Recall using the above TP, FN, and FP. Given that they are calculated from image pairs, we use the terms Improved Precision and Improved Recall in the following section.

C. Results

We evaluated the accuracy of our method and the following comparative methods.

- C_1: Comparative method 1
  In each image I_t and I_{t+1}, we applied the SSD [5] detector and predicted the candidate flower-bud regions. We trained the detector only on images in dataset 1 of acquisition condition 1. This method corresponds to applying only steps S1 and S3 of the proposed method in Fig. 1.
- C_2: Comparative method 2
  In each image I_t and I_{t+1}, we applied the YOLOv7-E6E[1] detector and predicted the candidate flower-bud regions. We trained the detector only on images in dataset 2

[1]https://github.com/WongKinYiu/yolov7/tree/main

(a) Dataset 1 (Acquisition condition 1)

(b) Dataset 2 (Acquisition condition 2)

Fig. 3. Examples of worm's-eye view images in our datasets.

(a) Flower buds in dataset 1 (b) Flower buds in dataset 2

Fig. 4. Examples of enlarged images of flower buds.

of acquisition condition 2. This method corresponds to applying only steps S1 and S3 of the proposed method in Fig. 1.

- C_3: Comparative method 3
  We applied the acquisition condition judgment described in Section II-B to select the suitable detector for each image. We prepared two detectors: SSD suitable for acquisition condition 1 and YOLOv7-E6E suitable for acquisition condition 2. We predicted the candidate flower-bud regions using the selected detector. This method corresponds to steps S1, S2, and S3 of our method in Fig. 1.
- O: Our method
  We applied the acquisition condition judgment described in Section II-B to each image and then applied the keypoint matching described in Section II-D to the image pairs. We prepared SSD and YOLOv7-E6E as detectors, as in the comparative method C_3. We used SuperPoint [10] and SuperGlue [11] for keypoint matching. This method corresponds to all steps of our method in Fig. 1.

TABLE I
ACCURACY OF COUNTING FLOWER BUDS USING EACH METHOD.

Method	Improved Precision ↑	Improved Recall ↑	Improved F-measure ↑
C_1: Comparative method 1	0.35	0.48	0.40
C_2: Comparative method 2	0.42	0.49	0.45
C_3: Comparative method 3	0.70	0.88	0.75
O: **Ours**	**0.81**	**0.88**	**0.84**

SSD was used for the comparative method C_1 and YOLOv7-E6E for the comparative method C_2 because these detectors achieved the highest detection accuracy for each dataset in the preliminary experiments. Training a single detector using images of both datasets did not improve the accuracy in preliminary experiments. SSD was trained with 500 images from dataset 1 and YOLOv7-E6E was trained with 500 images from dataset 2. The acquisition condition judgment was trained on the same images as those used to train the detectors. We completely separated images between the training and prediction processes. For evaluation of the prediction accuracy of counting flower buds, we prepared 85 image pairs from dataset 1 and 85 image pairs from dataset 2. The number of image segments was $B = 8$ for dataset 1 and $B = 6$ for dataset 2. The angle margin was $\theta_m = \pi/24$.

Table I shows the accuracy of counting flower buds using our method and the comparative methods. We used the Improved F-measure described in Section III-B. Our method O and the comparative method C_3 achieved higher counting accuracy than the comparative methods C_1 and C_2. This result showed that the acquisition condition judgment effectively solved the first issue, in which using a single detector at each time point decreases the count accuracy, described in Section I. Our method O achieved a higher count accuracy than the comparative method C_3. This result showed that keypoint matching effectively solved the second issue, in which duplicate counting between successive time points decreases the count accuracy, described in Section I. We confirmed that our method was capable of counting flower buds much more

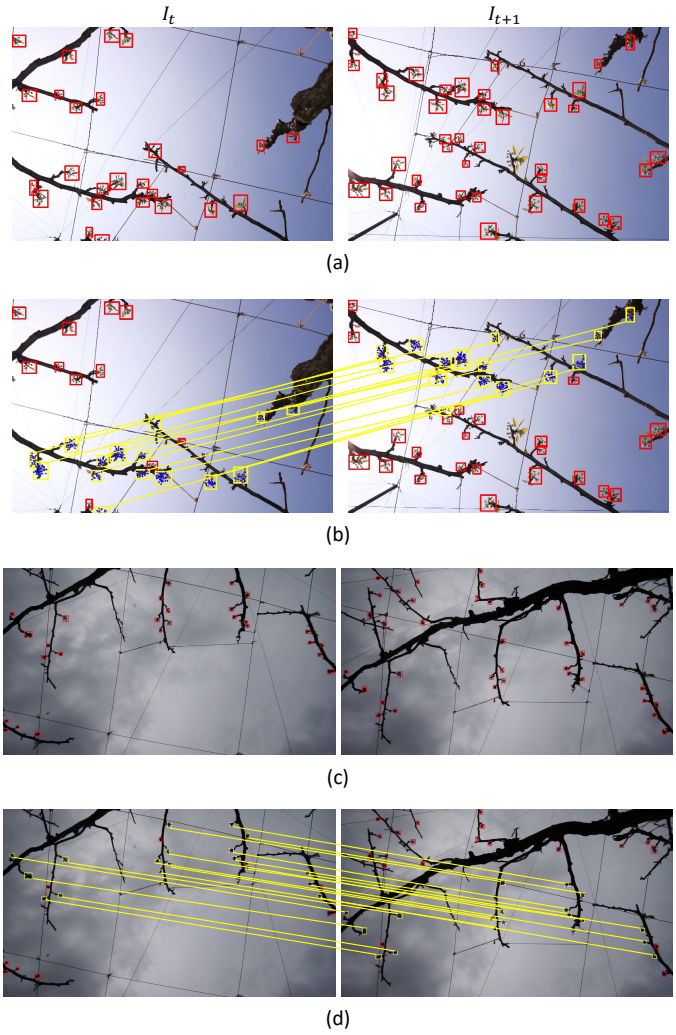

Fig. 5. Visualization of the results obtained using our method.

accurately than the comparative methods owing to the effects of the acquisition condition judgment and keypoint matching.

D. Visualization

Figure 5 shows examples of visualizing the results of counting pear flower buds using our method. In the figure, (a) and (b) are image pairs from dataset 1, and (c) and (d) are image pairs from dataset 2. In (a) and (c), the candidate flower-bud regions detected in step S3 of our method are represented by red boxes, and in (b) and (d), the keypoints of the candidate regions in step S4 are represented by blue points and the identical candidate regions by yellow boxes and yellow lines. We confirmed that flower buds were detected accurately by selecting a detector suitable for each acquisition condition, even when the apparent color of flower buds differed in each worm's-eye view image or when the size of the bounding box of the candidate flower bud changed. We also confirmed that candidate regions in overlapping view areas of image pairs were correctly determined, preventing duplicate counting of the identical flower buds.

IV. Conclusions

In this paper, we propose a method to detect candidate flower buds in worm's-eye view images at each time point, using a suitable detector selected based on the acquisition condition judgment and determining identical flower buds in an image pair using keypoint matching between successive time points. We confirmed that introduction of the acquisition condition judgment into our method improves the accuracy of counting flower buds when illumination and camera parameters vary, which is otherwise difficult to solve when using only a single detector at each time point. Furthermore, by introducing keypoint matching to our method, the accuracy of our method is improved when the camera system setting varies, which overcomes the risk of duplicate counting of identical flower buds between time points.

In future work, we intend to collect additional datasets for pear flower buds for evaluation of the counting accuracy by enriching the images of pear trees acquired under diverse acquisition conditions.

This work was partially supported by Research Center for Sustainable Science, Tottori University.

References

[1] K. I. Theron, "Size matters: Factors influencing fruit size in pear," in *Proceedings of the XI International Pear Symposium*, 2011, vol. 909, pp. 545–555.

[2] A. I. B. Parico and T. Ahamed, "Real time pear fruit detection and counting using yolov4 models and deep sort," *Sensors*, vol. 21, no. 14, 4803, 2021.

[3] N. Wouters, B. D. Ketelaere, J. D. Baerdemaeker, and W. Saeys, "Hyperspectral waveband selection for automatic detection of floral pear buds," *Precision Agriculture*, vol. 14, pp. 86–98, 2012.

[4] N. Wouters, B. D. Ketekaere, T. Deckers, J. D. Baerdemaeker, and W. Saeys, "Multispectral detection of floral buds for automated thinning of pear," *Computers and Electronics in Agriculture*, vol. 113, pp. 93–103, 2015.

[5] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. Berg, "SSD: Single shot multibox detector," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2016, pp. 21–37.

[6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788.

[7] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proceedings of the Advances in Neural Information Processing Systems (NIPS)*, 2015, vol. 28, pp. 91–99.

[8] G. Farjon, O. Krikeb, A. B. Hillel, and V. Alchanatis, "Detection and counting of flowers on apple trees for better chemical thinning decisions," *Precision Agriculture*, vol. 21, pp. 503–521, 2020.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[10] D. DeTone, T. Malisiewicz, and A. Rabinovich, "SuperPoint: Self-supervised interest point detection and description," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018, pp. 337–33712.

[11] P. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich, "SuperGlue: Learning feature matching with graph neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 4937–4946.

[12] T. Hanakawa, M. Nishiyama, Y. Iwai, E. Morimoto, and S. Nakatani, "Counting pear flower buds using worm's-eye view images," in *Proceedings of IEEE 11th Global Conference on Consumer Electronics (GCCE)*, 2022, pp. 544–547.