Automatic position correction using center estimation for cereal images

Kotarou Saku*, Takumi Ozaki*, Shun Sakuma†, Michiko Inoue* and Masashi Nishiyama*,
*Graduate School of Sustainability Science, Tottori University, Japan
†Faculty of Agriculture, Tottori University, Japan

***Abstract*—When acquiring images to construct a database of the appearance characteristics of cereal species, we need to be aware of the misalignment depending on how each individual cereal spike is placed and how the scanner is used. In this paper, we propose a method to automatically correct the misalignment by estimating the center position of a set of spikelets from a cereal image using a class activation map of cereal species classification. We demonstrate that our method can accurately estimate the center position of a set of spikelets and correct the misalignment of cereal images.**

***Index Terms*—Cereals, Position correction, Center estimation**

I. Introduction

There is a significant demand for superior cereal crops, such as barley and wheat, to provide a stable food supply for the growing population. As part of this effort, a genetic database[1] is being developed to store and publish data for various cereal species. Recent attempts have been made to store the appearance characteristics of cereal species, in addition to their genetic data, in the new database. This database stores data pertaining to cereal genes, growing conditions (year, field, weather), and individual appearance characteristics (i.e., images of various parts of cereal plants). Expert agronomists hope to discover new appearance characteristics of cereal species by analyzing images stored in the database with reference to genetic data and growing conditions. Figure 1(a) shows an example of an image of a cereal spike obtained by scanning. A cereal spike comprises a set of spikelets, an awn, and a stem. A spikelet is an individual component of a cereal spike, and we refer to the dense region where they are clustered as a set of spikelets.

To collect cereal images for storage in the database, each cereal spike has to be photographed individually, which is a significant effort when acquiring many cereal images. To reduce the effort required to acquire these images, it would be useful to enlist the help of non-experts and use an inexpensive scanner to acquire cereal images in a small space. Specifically, to acquire a cereal image, a non-expert places the individual cereal spike on the scanner stand, closes the scanner cover, and scans it. However, we need to be aware that each acquired image is misaligned to some degree, depending on how the individual cereal spike is placed and how the scanner cover is closed. Figure 1(b) shows an example of a misaligned cereal image. In this example, the center position of the set of spikelets of each individual cereal spike is shifted vertically and horizontally in the cereal image. This misalignment causes problems when expert agronomists analyze the appearance characteristics of cereal species based on cereal images, because they focus on a set of spikelets in a cereal spike. The center position of the set of spikelets should be constant without shifting among all of the cereal images stored in the database.

[1] https://wheat.pw.usda.gov/GG3/

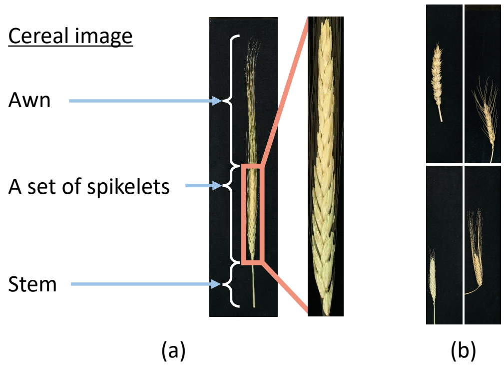

Fig. 1. Examples of cereal images acquired from cereal spikes, such as barley and wheat.

We propose a method to automatically correct the misalignment by estimating the center position of a set of spikelets from a cereal image using a class activation map (CAM). This method is useful for correcting cereal images acquired by non-experts for storage in databases. We experimentally confirmed that the CAM responds near a set of spikelets in a deep neural network that classifies cereal species. This resembles the analytical strategy of expert agronomists as they focus on a set of spikelets when distinguishing the appearance characteristics of cereal species. Thus, our method uses the CAM of the cereal species classification to estimate the center position of the set of spikelets and correct the misalignment of a given cereal image. Our experiments confirmed that our method can accurately estimate the center position of a set of spikelets and correct the misalignment of cereal images.

II. Our Method

A. Limitations of center estimation by simple segmentation

Before describing our method, we explain segmentation, which is one of the methods used for correcting misalignment. Segmentation estimates the region of a set of spikelets from

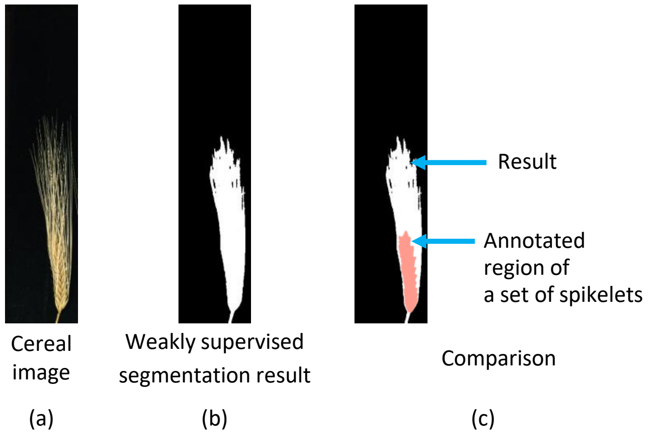

Fig. 2. Comparison of an estimation result from weakly supervised segmentation with an annotated region of a set of spikelets.

the cereal image and determines its center position. Generally, segmentation methods [1], [2] perform fully supervised learning, which requires the preparation of region masks that are manually annotated in pixel units of the spikelets. The preparation for fully supervised learning incurs significant costs. Recently, weakly supervised segmentation methods [3], [4] have been proposed that use only class labels as the supervised signal. The advantage is that a region mask is not needed for each cereal image; only the cereal species name needs to be assigned manually. We note, however, that weakly supervised segmentation does not have a mechanism for segmenting only the region of a set of spikelets, so the segmentation results may include unnecessary regions. For example, we applied WeakTr [4], one of the weakly supervised segmentation methods, to the cereal image in Fig. 2(a), and obtained the segmentation result shown in (b). When this segmentation result was superimposed on the manually annotated region of a set of spikelets, we observed that it contained many unnecessary regions, as shown in (c). Ideally, the center position should be estimated from the region of the set of spikelets. If the center position is estimated from the segmentation results, including the unnecessary regions, then the estimation error of the center position becomes larger.

B. Center estimation and position correction using the CAM of cereal species classification

To correct the misalignment of cereal images, we aimed to obtain a candidate region that included the region of a set of spikelets but did not include unnecessary regions. Expert agronomists focus on a set of spikelets when distinguishing the appearance characteristics of cereal species. We confirmed that the CAM also responds near a set of spikelets in deep neural networks for classifying cereal species. On the basis of these findings, we aimed to narrow down the candidate region for a set of spikelets. We note, however, that unnecessary regions were also included in the CAM. To improve the accuracy of center estimation, we used a supplementary map representing the entire cereal spike obtained using binarization of the cereal image. We aimed to remove unnecessary background regions from the regions obtained from the CAM.

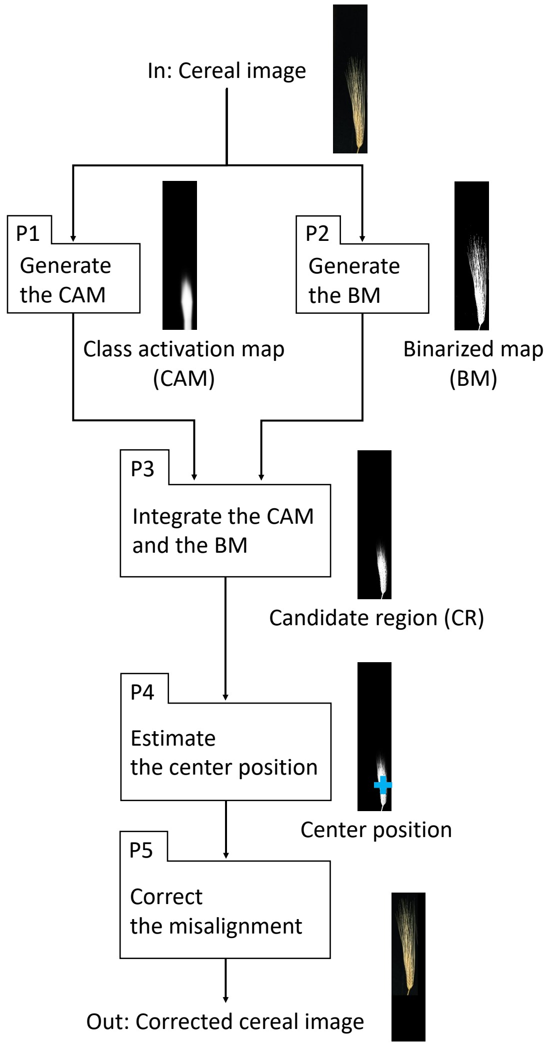

Fig. 3. Overview of our method.

Figure 3 shows the overview of our method. After acquiring a scanned cereal image, we used our method to estimate the center position and correct the misalignment as follows:

P1: We generated the CAM from the cereal image using a deep neural network for cereal species classification. In addition, our method applied a highlighting process to the CAM. Specifically, the values of each pixel on the CAM were normalized between 0 and 1, and the following equation was computed:

$$y = \frac{1}{1 + e^{-a(x-0.5)}} \tag{1}$$

where x is the value of each pixel in the CAM and y is the value after the highlighting process. $a(> 0)$ is a parameter that controls the degree of emphasis.

P2: We then generated a binarized map (BM) representing the whole body of the cereal spike from the cereal image. The background of the scanned cereal image was assumed to be black. Our method estimated a map representing the whole body by binarization using Otsu's method.

P3: We determined the candidate region (CR) by integrating the CAM obtained in P1 and the BM obtained in P2 to remove unnecessary regions, such as the background. Specifically, the CR was determined by finding the Adamar product on each pixel between the CAM and the BM.

P4: To estimate the center position of the set of spikelets, we computed the center of gravity of the CR obtained in P3.

P5: Finally, we corrected the misalignment by translating the cereal image using the center position obtained in P4.

III. Experiments

A. Dataset

To confirm the effectiveness of our method, we collected cereal images of the following four cereal species.

- Barley
- Einkorn wheat
- Durum wheat
- Bread wheat

We prepared 153 barley, 153 einkorn wheat, 153 durum wheat, and 153 bread wheat spikes. One cereal image per individual spike was acquired using a scanner (GT-X830, Seiko Epson Co., Nagano, Japan). A total of 612 cereal images were collected. The size of each cereal image was set to 800×182 pixels. To evaluate the accuracy of the center estimation, we manually annotated the region of a set of spikelets in each of the 612 images. We used the center of gravity of the annotated region as the correct value of the center position. To train the deep neural network for species classification, we prepared a total of 2,892 training samples for individual spikes other than those in the above images: 723 for barley, 723 for einkorn wheat, 723 for durum wheat, and 723 for bread wheat.

B. Experimental conditions

Our method used ResNeXt [5] to classify cereal species and LayerCAM [6] to generate the CAM. Specifically, we applied LayerCAM to the third Bottleneck Block of conv5 of ResNeXt. The highlighting parameter of Eq. (1) in P1 was set to $a = 16$.

To evaluate the accuracy of center estimation, we used the error of the center position of the set of spikelets. Specifically, the Euclidean distance was calculated between the correct value of the center position described in Section III-A and the value estimated using our method. We calculated the error for each of the 612 cereal images with the manually annotated region of the set of spikelets and calculated the mean and standard deviation.

C. Results of the center estimation and position correction

We evaluated the estimation error of the center position of the set of spikelets using the following methods.

- Baseline: Cereal image acquired by a non-expert with no further processing.
- Weakly supervised segmentation: Image processed using WeakTr [4] to find the region of the set of spikelets and estimate the center position using the center of gravity of the segmentation result.
- Our method: Image processed to estimate the center position on the basis of the CAM and the BM using our method described in Section II-B,

TABLE I
ESTIMATION ERROR OF THE CENTER POSITION OF A SET OF SPIKELETS.

Method	Error [pixels]
Baseline	178±54
Weakly supervised segmentation	44±32
Ours	**29±21**

Table I shows the estimation error of the center position of the set of spikelets. Among all the tested methods, our method had the smallest estimation error, 29±21 pixels. We applied the Wilcoxon signed rank sum test and Bonferroni correction at a significance level of $p < 0.05$. The results revealed significant differences among the methods. These results confirmed that our method was superior in estimating the center position of the set of spikelets in the cereal images. Figure 4(a) shows cereal images acquired by non-experts, and (b) shows the cereal images after center position correction. Compared with the non-processed images acquired by non-experts (i.e., images without center estimation and position correction), the cereal images corrected by our method showed better alignment of the set of spikelets.

D. Effectiveness of the CAM and the BM

Next, we evaluated the effectiveness of using the BM in addition to the CAM. Here, we compared the estimation error of the center position for the following methods:

- Comparative method 1: We estimated the center position using only the CAM. Specifically, we processed P1, P4, and P5 in Fig. 3 in this order.
- Comparative method 2: We estimated the center position using only the BM. Specifically, we processed P2, P4, and P5 in Fig. 3 in this order.
- Our method: We estimated the center position using the CAM and the BM. Specifically, we processed P1, P2, P3, P4, and P5 in Fig. 3 in this order.

All other experimental conditions were the same as those described in Section III-C.

Table II shows the estimation error when using the CAM and the BM. The estimation error was the smallest when our method was used. We applied the Wilcoxon signed rank sum test and Bonferroni correction at a significance level of $p < 0.05$. The results revealed significant differences among the methods. This confirmed the effectiveness of using a BM as an adjunct to the CAM.

Figure 5 shows examples of the CAM and the BM generated in our method. The CAM obtained from each cereal image (WI) contained unnecessary regions, such as the background, when compared with the manually annotated region of the

Fig. 4. Comparison of corrected cereal images. (a) Cereal images acquired by non-experts with no further processing; (b) cereal images after performing center estimation and position correction.

TABLE II
ESTIMATION ERROR WHEN USING THE CAM AND THE BM

Method	Map	Error [pixels]
C1	Class activation map (CAM)	38±26
C2	Binarized map (BM)	66±31
Ours	**CAM and BM**	**29±21**

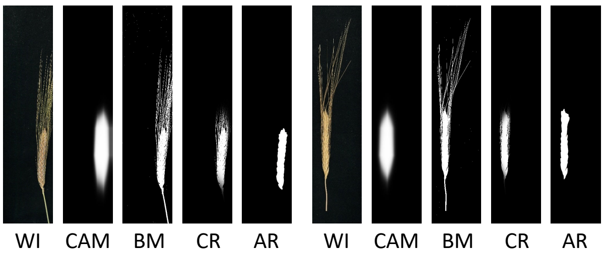

Fig. 5. Examples of the CAM and the BM generated using our method. From left to right: cereal image (WI), class activation map (CAM), binarized map (BM), candidate region (CR), and annotated region (AR) of a set of spikelets.

set of spikelets. Like the CAM, the BM also contained many unnecessary regions. In contrast, the candidate region (CR) generated by our method was similar to the manually annotated region of the set of spikelets.

IV. Conclusions

We propose a method to automatically correct the misalignment of cereal images acquired by non-experts. This method will be useful for processing images before storage in databases that include the appearance characteristics of cereal species. Our method estimates the center position of the set of spikelets from a cereal image using the CAM of the cereal species classification in addition to a BM. Our experimental results demonstrate the effectiveness of our method in correcting the misalignment in cereal images acquired by non-experts, thereby instilling confidence in its practical application.

In future work, we intend to develop methods to further reduce the workload of non-experts when acquiring cereal images, and a method to cope with the rotational shift of cereal spikes.

References

[1] Z. Zhou, M.R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pp. 3–11. 2018.

[2] L. Chen, G.Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 04, pp. 834–848, 2018.

[3] Y. Du, Z. Fu, Q. Liu, and Y. Wang, "Weakly supervised semantic segmentation by pixel-to-prototype contrast," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 4320–4329.

[4] Lianghui Zhu, Yingyue Li, Jiemin Fang, Yan Liu, Hao Xin, Wenyu Liu, and Xinggang Wang, "Weaktr: Exploring plain vision transformer for weakly-supervised semantic segmentation," in *arXiv*, 2023.

[5] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He, "Aggregated residual transformations for deep neural networks," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5987–5995.

[6] Peng-Tao Jiang, Chang-Bin Zhang, Qibin Hou, Ming-Ming Cheng, and Yunchao Wei, "Layercam: Exploring hierarchical class activation maps for localization," *IEEE Transactions on Image Processing*, vol. 30, pp. 5875–5888, 2021.