Development of AI-based System for Classification of Objects in Farms Using Deep Learning by Chainer and a Template-Matching Based Detection Method

Shinji Kawakura
Laboratory for Future Interdisciplinary Research in Science and Technology, Tokyo Institute of Technology, Kanagawa, Japan
Email: kawakura.s.aa@m.titech.ac.jp

Ryosuke Shibasaki
Center for Spatial Information Science, The University of Tokyo, Meguro, Japan
Email: shiba@csis.u-tokyo.ac.jp

***Abstract*—It has generally been difficult for agri-system developers to identify diverse objects automatically and accurately before the harvesting without touching something dangerous (e.g., poisonous creatures, toxic substances). Such objects could include harvestings for sale, stems, leaves, artificial stiff frames, unnecessary weeds, agri-tools, and creatures, especially in Japanese traditional small-medium sized, insufficiently trimmed (messed) farmlands. Scientists, agri-managers, and workers have been trying to solve these problems. On the other side, researchers have been advancing robot systems, mainly based on automatic machines for harvesting and pulling up weeds utilizing visual-data analysis systems. These studies have captured a significant amount of visual data, identified objects with short time delay. However, previous products have not yet met these requirements. We have considered the achievements of recent technologies to develop and test new systems. The purpose of this research is proving the utility of this visual-data analysis system by classifying and outputting datasets from an AI-based image system that obtained field pictures in outdoor farmlands. We then apply Chainer for deep learning, and focus on computing methodologies relating to template-matching and deep learning to classify the captured objects. The presented sets of results confirm the utility of the methodologies to some extent.**

***Index Terms*—classification of objects, chainer, deep learning, identifying outdoor things using template-matching based method, OpenCV**

I. Introduction

Several agri-systems for diverse farmlands have been developed, which have, in some way, reached a good level for several informatics and mechanics researchers. Many techniques have also been developed to enhance the utility of multiple robot systems to analyze and identify various objects, including vegetables and fruits [1]-[20]. Past achievements in several related fields have been acknowledged across the world. Additionally, researchers in the field of agricultural informatics and robotics continue to propose promising methods for improving these tasks. However, existing visual analysis methods, including robotic farming systems, have focused mainly on vegetables and fruits for sale, scattering weeds, artificial stiff frames used for stability, and humans (farmers), and have given insufficient attention to peripheral objects [1], [10], [11], [13], [15], [16], [21]. Some past studies have analyzed such objects using their colors and textures.

In this study, we created agriculture datasets to develop a visual data analysis system based on novel deep learning using Chainer (Preferred Networks Inc., Japan) and template matching methodologies. We performed analyses using a combination of our programs and distributed Chainer, OpenCV libraries and others. In contrast to other countries vast agri-fields (e.g., large-sized farmlands in the United States, European countries, and China), the target environments of this study were traditional Japanese small-to-medium-sized fields with insufficiently trimmed (messed) farmlands. We also targeted the following objects: mini tomatoes and kiwi fruits (for sale), mini tomato stems and leaves, and (unnecessary) weeds.

II. Materials and Methods

After investigating the following freely distributed and sophisticated frameworks for deep learning-based computing: Chainer, Caffe, Caffe2, TensorFlo, Kera, Torch, PyTorch, and MatConvNet, we chose Chainer, which is a layer-oriented deep learning framework that can be used to analyze picture datasets.

The benefits of Chainer include its ability to perform complicated computing calculations, its coverage across

Manuscript received April 3, 2019; revised August 21, 2019.

doi: 10.18178/joaat.6.3.175-179

simple network-systems to relatively complex deep learning flexibly, high speed computing using a graphics processing unit (GPU), and its description of deep learning systems in Python. Its disadvantages are the slower calculation speed than programs written in other recent languages. Further, we made programs for the visual template matching trials including libraries and header files of OpenCV ver.3.2 for Visual Studio ver. 2015 and 2017.

A. *Target*

First, we carefully categorized the following small creatures in agri-fields by their physical and visual characteristics (color, shape, size, etc.).

Based on feedback from real farmers, this study targeted (1) mini tomatoes (training data set: $n_1 = 5$, analyzed data set: $n_2 = 16$); (2) kiwi fruits ($n_1 = 5$, $n_2 = 13$), ((1) and (2) are for sale); (3) mini tomato stems ($n_1 = 5$, $n_2 = 13$); (4) mini tomato leaves ($n_1 = 5$, $n_2 = 13$); and (5) unnecessary weeds ($n_1 = 5$, $n_2 = 16$), as shown in Fig. 1. Each picture was standardized into 56 × 56 pixels.

For the capturing module, to ensure future multi-purpose evolutions of the system, we selected and used a non-specific (not for academic) digital camera (Cybershot DSC-WX300, Sony Inc., Japan). Evolutions.

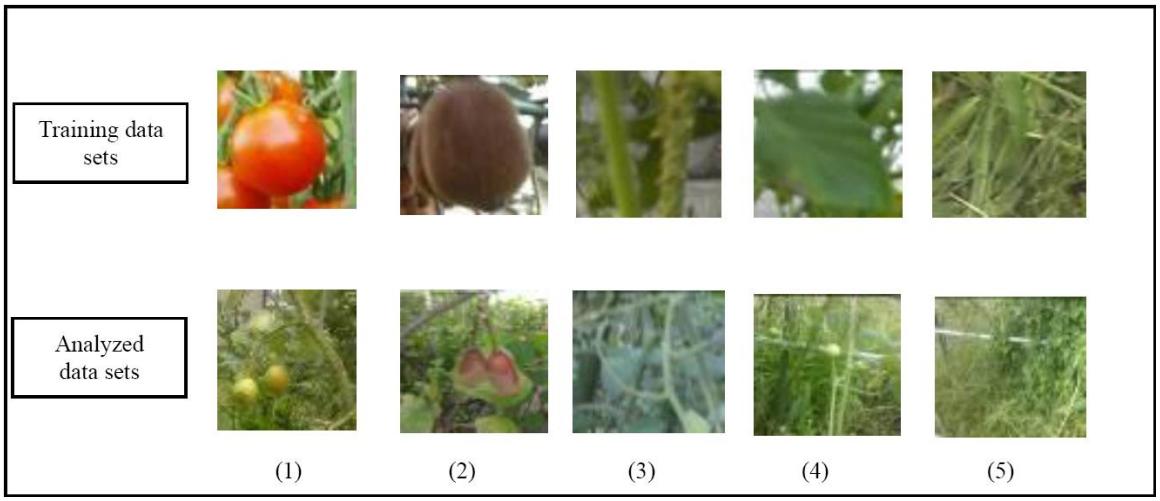

Figure 1. Example sets pictures from each data set concerning the target for each target: (1) mini tomato, (2) kiwi fruit, (3) mini tomato stem, (4) mini tomato leaf, and (5) weeds.

B. *Objective of the Analyses*

We created original training datasets without using recent distributed open datasets as the target or training data (reference data). In current computer science fields, some studies have effectively used these open datasets. Their methodology has adequate universality in the agricultural informatics field. Fig. 2 shows the flow of the experiment, which comprised (1) setting up the necessary computing systems, (2) capturing and accumulating picture data from outdoor farmlands, (3) creating a training dataset using Chainer for each target object, (4) classifying other sample picture datasets to select the most compatible/plausible training picture, (5) testing the method of template matching and obtaining the resulting data, and (6) contributing the results to recent, high-tech agricultural systems developers.

Fig. 3 shows the flow of the program written in Python for the computing processes to make the training data sets. It shows our system for the Chainer Framework Framework-based analyses, which comprised multiple layers. For the template-matching's sequential processing, we programmed and used the system written in Visual C++ in the environment of Visual Studio 2015 and 2017 including OpenCV header files and libraries.

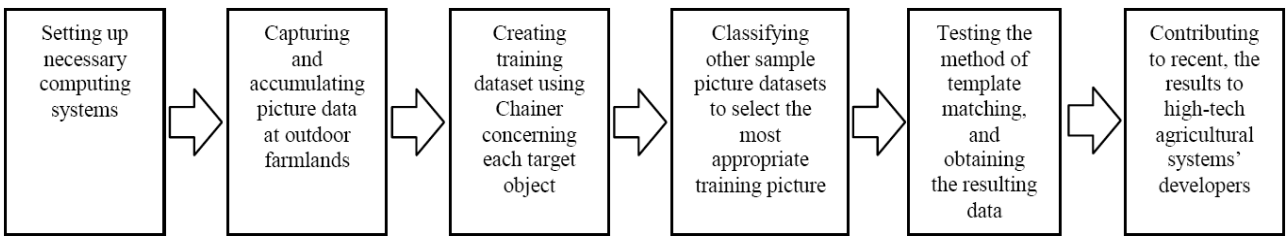

Figure 2. Flow of the experiment.

[Pre-processing]
・Initial imports of library files from Chainer and small functions in included in Chainer file set
・Declaration of small functions and links of Chainer

[Training model setting function using multi-layers' neural network (A)]
・Initializing variables
・Calculating and returning numerical data for neural network

[Changing picture data sets function (B)]
・Changing pictures' size and formats for picture recognition
・Standardizing pixel values of all pictures
・Transforming the size of matrix of the pixels from (56 px ×56 px) into (1 px ×56^2 px)
・Returning the matrix data

[Main function]
・Calling function (A)
・Including raw data pictures

[While loop]
・Cutting picture files into rectangle forms
・Representing the pictures
・Calling function (B)
・Determining the most appropriate training pictures' number

[Pushing indicated button to finish]
・Escaping from the loop
・Releasing capture data
・Closing all windows

[Post-processing]
・Waiting and closing the system.

Figure 3. Flowchart of Chainer-based deep learning program written in Python.

III. RESULTS

In Fig. 4, we present a sample output: five-dimensional probability vectors shown in a console window using a Chainer-based deep learning program written in Python. Table I provides data from the statistical analyses for the set of objects. In the figure, the column of the vector with the highest data value indicates the most compatible sample picture data, which is "#2" shown in the figure.

```
. . .

[ 1.17279704, 3.24628832, -1.08857703, -2.48619319, 1.98607193]

The most compatible sample number
#2

[ -0.29490658, 1.85294889, 2.95399467, 1.56906870, -0.94809723]

The most compatible sample number
#3

. . .
```

Figure. 4 Sample analyzed data (five-dimensional probability vectors) output in console window.

Fig. 5 presents a sample output concerning our visual data template-matching process with larger pictures than presented in similar past studies. In previous phases, we determined the most compatible picture data (see Table I). In this phase, we used these data in the template matching based analyses. Fig. 5 shows the most appropriate (closest to the average according to the brightness and deviation of the scatterings) sample.

Figures 5. Sample picture data from the template matching analysis for a mini-tomato farm. (In the right picture, some highlighted parts have higher possibilities that the area is similar to the selected training picture, shown in the left.)

TABLE I. NUMBERS OF MOST MATCHED (COMPATIBLE) SAMPLE PICTURE DATA: WE COMPARED 5 TRAINING DATA PICTURES (N = 5, #1 - #5) SELECTED BY MATURE WORKERS WITH A SAMPLE DATA SET USING THE CHAINER-BASED DEEP LEARNING PROGRAM. (FOR INSTANCE, CONCERNING "(1) MINI TOMATO," #3 HAD THE MOST MATCHED DATA, SO WE USED TRAINING DATA #3 FOR THE SUCCESSIVE TEMPLATE MATCHING ANALYSES)

Item			Number of most matched (compatible) sample picture data				
Training data			# 1	# 2	# 3	# 4	# 5
(1)	Mini tomatoes	(n = 16)	3	2	6	1	4
(2)	Kiwi fruits	(n = 13)	4	0	3	3	3
(3)	Mini tomato stems	(n = 13)	6	3	4	0	0
(4)	Mini tomato leaves	(n = 13)	5	4	1	2	1
(5)	Unnecessary weeds	(n = 16)	1	1	11	3	0

IV. DISCUSSION

For various future contributions to agri-systems developers, we obtained significant output for the evolution of future systems and confirmed the usefulness of deep learning-based and template matching-based approaches. As presented in Table I, for the cases of the mini-tomatoes, for instance, #3 in the dataset had the highest matching value. It is difficult to comment on the cases concerning Table I and Fig. 5 because they were just trial experiments, and their colors and textures are similar to various objects behind them. However, for both visual examinations (inspections) and the template-matching data, obvious differences were evident between the mini-tomatoes and other green and brown objects. We were able to confirm these tendencies from the analyzed numerical data and captured picture data. We judged the highlighted positions to be certain to some extent. In contrast to the responses from interviews with real agri-workers and managers, the numbers in the data in Table I regarding the most matched (compatible) sample picture data became rather inhomogeneous.

Additionally, we should rethink the sample number (Training data: 5, and (1)-(5): 13–16), because, in many past studies [3], [5], [6], [8], [9], [11], [12], [14], [16], [21] researchers and engineers used several hundred or thousand samples. The data of "(3) Mini tomato stem" and "(5) Unnecessary weeds" seemed appropriate for this study thinking of the aforementioned past similar studies. Overall, we found that the most accurate target object from the template-matching analyses was the (1) mini tomato. Based on the trials of the most compatible training data samples, we consider that these systems would be useful for future studies with increased sample volume and object varieties.

In later studies, we should improve their accuracy.

V. CONCLUSION AND FUTURE TASKS

W We constructed and demonstrated a picture data set categorizing system utilizing a deep learning method based on the Chainer framework. We also executed template-matching-based trials, selecting appropriate training picture data and widely capturing outdoor farmland pictures. Various applied systems should be tested using such methodologies at real agricultural sites. This study demonstrated the system's accuracy and ability to some extent and presented various numerical datasets. The captured picture files were judged by mature agri-workers, and the files were analyzed automatically using the system's classification with changing picture types and conditions (e.g., mini-tomatoes were ripe, distances between targets and camera), with consideration for future practical usage. In addition, we described the differences between two pattern data groups. Considering the history of visual data based deep learnings, we should not and cannot simply compare these findings to past similar fields' academic achievements because of the insufficiency of the volume and quality of appreciated samples in agricultural fields. In light of recent high-tech agricultural techniques and different sized farmlands, we confirmed the usefulness and scalability of our proposed methodologies. In our future work, we aim to provide further confirmation of the validity, durability, and performance of the system for a variety of detected targets and background conditions. This may include methodologies for automatic agricultural machinery, such as cultivating systems, weeding machines, and cutting vegetables. Patterns from other databases might also be helpful in such fields. We hope the aforementioned promising methodologies will be widely applied to real working sites to promote the recruitment of workers into agricultural fields. These methods could also prevent farmers from various accidents (toxic substances, dangerous small species, etc.). Finally, in the long term, these results could be used for automatic systems to help indoor farmers (workers in greenhouses, etc.) and outdoor farmers revise and improve their work practices.

REFERENCES

[1] S. S. Mehta and T. F. Burks, "Vision-based control of robotic manipulator for citrus harvesting," *Computers and Electronics in Agriculture*, vol. 102, pp. 146-158, January 2014.

[2] A. Shinjo and M. Kudo, "The practical use of IT in agriculture: The movement into high-value-added crops and integrated solutions," *The Journal of the Institute of Electronics, Information, and Communication Engineers*, vol. 96, no. 4, pp. 280-285, April 2013.

[3] Y. Huang, "The advancement of nature-inspired algorithms for agriculture," in *Proc. American Society of Agricultural and Biological Engineers (ASABE) Annual International Meeting*, 2018, p. 1.

[4] N. Zhu, X. Liu, Z. Liu, K. Hu, Y. Wang, J. Tan, and Y. Guo, "Deep learning for smart agriculture: Concepts, tools, applications,

and opportunities," *International Journal of Agricultural and Biological Engineering*, vol. 11, no. 4, pp. 32-44, July 2018.

[5] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, "Deep neural networks based recognition of plant diseases by leaf image classification," *Computational Intelligence and Neuroscience*, pp. 1-12, May 2016.

[6] F. A. Guth, S. Ward, and K. P. McDonnell, "Autonomous disease detection in crops using deep learning," *Biosystems and Food Engineering Research Review*, vol. 22, pp. 1-209, May 2017.

[7] K. Ahlin, B. Joffe, A. P. Hu, G. McMurray, and N. Sadegh, "Autonomous leaf picking using deep learning and visual-servoing," *IFAC-PapersOnLine*, vol. 49, no. 16, pp. 177-183, October 2016.

[8] E. Cengil, A. Çınar, and E. Özbay, "Image classification with caffe deep learning framework," in *Proc. IEEE International Conference in Computer Science and Engineering*, 2017, pp. 440-444.

[9] S. W. Chen, S. S. Shivakumar, S. Dcunha, J. Das, E. Okon, C. Qu, and V. Kumar, "Counting apples and oranges with deep learning: A data-driven approach," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 781-788, December 2016.

[10] K. Ahlin, B. Joffe, A. P. Hu, G. McMurray, and N. Sadegh, "Autonomous leaf picking using deep learning and visual-servoing," *IFAC-Papers on Line*, vol. 49, no. 16, pp. 177-183, October 2016.

[11] T. T. Zin, C. N. Phyo, P. Tin, H. Hama, and I. Kobayashi, "Image technology based cow identification system using deep learning," in *Proc. International Multi Conference of Engineers and Computer Scientists*, vol. 1, pp. 320-323, March 2018.

[12] M. Brahimi, K. Boukhalfa, and A. Moussaoui, "Deep learning for tomato diseases: Classification and symptoms visualization," *Applied Artificial Intelligence*, vol. 31, no. 4, pp. 299-315, April 2017.

[13] A. Dutta, J. M. Gitahi, P. Ghimire, and R. Mink, "Weed detection in close-range imagery of agricultural fields using neural networks," *Publikationen der DGPF, Band 27*, pp. 633-645, 2018.

[14] F. J. Rodríguez, A. García, P. J. Pardo, F. Chávez, and R. M. Luque-Baena, "Study and classification of plum varieties using image analysis and deep learning techniques," *Progress in Artificial Intelligence*, vol. 7, no. 2, pp. 119-127, January 2018.

[15] E. Okafor, R. Smit, L. Schomaker, and M. Wiering, "Operational data augmentation in classifying single aerial images of animals," in *Proc. IEEE International Conference on Intelligent Systems and Applications*, 2017, pp. 354-360.

[16] R. Wang, J. Zhang, W. Dong, J. Yu, C. J. Xie, R. Li, and H. Chen, "A crop pests image classification algorithm based on deep convolutional neural network," *Telkomnika*, vol. 15, no. 3, pp. 1239-1246, August 2017.

[17] M. Beyeler, *Machine Learning for OpenCV Intelligent Image Processing with Python*, The United States: CUTT SYSTEM, 2018.

[18] N. Kitayama, *OpenCV3 Kihon Programming*, Tokyo, Japan: CUTT SYSTEM, 2017.

[19] N. Kitayama, *Python de Hajimeru OpenCV3 Programming*, Tokyo, Japan: CUTT SYSTEM, 2016.

[20] N. Toyama and M. Tsujitani, *Jissen R Toukei Bunseki*, Tokyo, Japan: Ohmsha, 2015.

[21] N. Nagata and S. Toyosawa, *Jissen OpenCV 3 for C++*, Tokyo, Japan: CUTT SYSTEM, 2017.

Shinji Kawakura received Ph.D. in Environmentology from the University of Tokyo in 2015, in Meguro-ku, Tokyo, Japan.
He was born on July 14, 1978. He got B.A in control system engineering from Tokyo Institute of Technology in 2003; M.A. in human-factors engineering from Tokyo Institute of Technology in 2005 in Meguro-ku, Tokyo, Japan. He worked on system engineering, researching for private companies. His interests include Development and Verification of Wearable Sensing Systems with Real-time Spoken Commands for Outdoor Agricultural Workers.

Ryosuke Shibasaki is with Department of Socio-Cultural and Socio-Physical Environmental Studies, The University of Tokyo/Kashiwa-shi, Chiba, Japan. He is Dr. in Engineering. Now he is Professor at the Center for Spatial Information Science, University of Tokyo.