Automatic Flower and Visitor Detection System

Dat Thanh Tran*, Toke Thomas Høye†, Moncef Gabbouj*, Alexandros Iosifidis‡
*Laboratory of Signal Processing, Tampere University of Technology, Finland
†Department of Bioscience, Aarhus University, Aarhus, Denmark
‡Department of Engineering, Electrical & Computer Engineering, Aarhus University, Denmark
Email: {dat.tranthanh,moncef.gabbouj}@tut.fi, tth@bios.au.dk, alexandros.iosifidis@eng.au.dk

***Abstract*—The visit patterns of insects to specific flowers at specific times during the diurnal cycle and across the season play important roles in pollination biology. Thus, the ability to automatically detect flowers and visitors occurring in video sequences greatly reduces the manual human efforts needed to collect such data. Data-dependent approaches, such as supervised machine learning algorithms, have become the core component in several automation systems. In this paper, we describe a flower and visitor detection system using deep Convolutional Neural Networks (CNN). Experiments conducted in image sequences collected during field work in Greenland during June-July 2017 indicate that the system is robust to different shading and illumination conditions, inherent in the images collected in the outdoor environments.**

I. Introduction

In recent years, deep CNNs have become the core component in several visual inference tasks ranging from object classification, object detection to saliency segmentation [1], [2], [3], [4], [5]. Over the years, the development of Graphical Processing Units (GPUs) and the availability of large-scale datasets have enabled us to train large and deep CNNs with millions of parameters with better and better generalization performance. The success of CNNs in recognizing human-familiar objects such as those in ImageNet challenge has motivated researchers in different fields to validate the use of CNNs in problems of their domain. While the performance of an automatic visual inference system depends on several factors, such as the amount of available data, the resolution of the input images or the variation in poses, shading or illumination conditions, CNNs have shown promising results in several application domains [6], [7], [8], [9], [10], [11].

In order to study the behaviors of different types of insects that visit a flower bed and their impact on pollination or reproduction patterns, several types of information should be analyzed, including visitation rates or visiting patterns between flowers within the same field. Data collection following a manual process greatly based on human efforts is traditionally done by observations in the field [12], or more recently by searching and annotating frames of video sequences recorded in the field [13]. Without any automation, data collection step is costly and inherently time-consuming, since it accounts for long time periods to collect adequate information to draw reliable conclusions. For example, by observing the visiting patterns of insects throughout four-month periods, the authors in [14] noticed the high discrepancy in frequency between different types of insects. With the availability of an automatic system that can reliably detect flowers within a video frame and further perform visitor detection, behavioral patterns of individual flowers and visitors can be captured, helping in providing extensive statistical evidence, validation or insights into the pollination process or the response of both plants and insects to environmental changes.

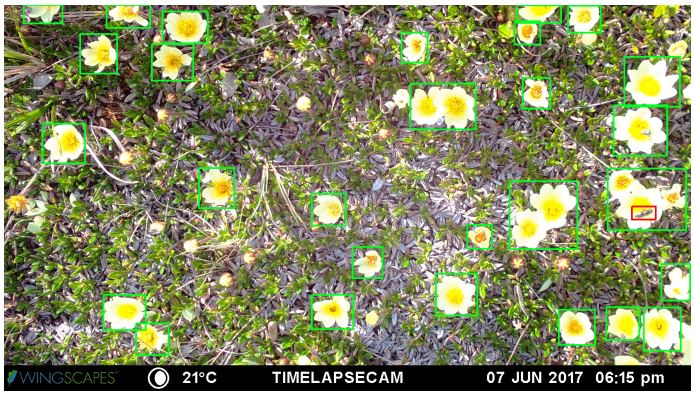

Fig. 1. Bounding box annotations for flowers (green) and visitor (red)

While using an automatic visual inference system seems to be straightforward, there are inherent factors that could affect performance in practice, such as different imaging conditions throughout the day. To the best of our knowledge, there is yet any work that attempts to validate traditional image processing techniques, such as image segmentation, or the use of machine learning models to build a flower and visitor detection system in the wild, that would facilitate the study of related biological indices. In this paper, we describe a method for both flower and visitor detection in unconstrained conditions (*in the wild*). During a preliminary study, we identified weaknesses of standard image processing approaches related to different lighting conditions and light exposure changes that might appear in the real application scenario, leading to unreliable performance levels. To overcome these issues, we propose a Machine Learning solution that exploits CNNs for both flower-based image segmentation and visitor detection.

ISBN 978-90-827970-1-5 © EURASIP 2018

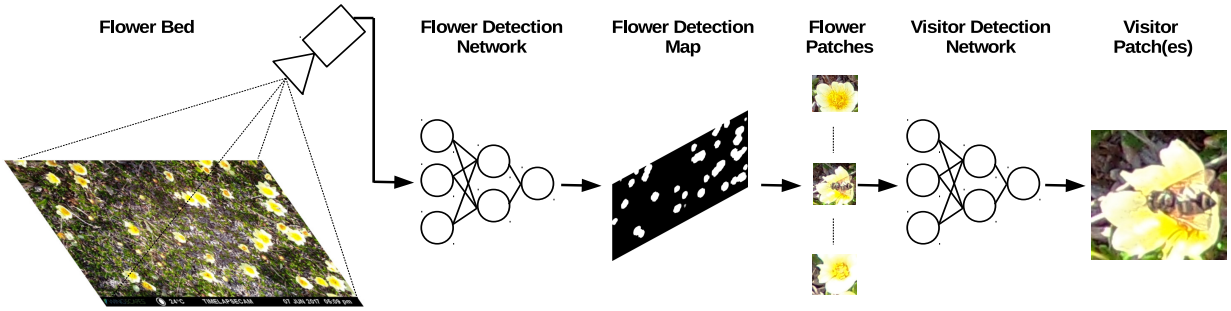

Fig. 2. Flower and visitor detection pipeline

II. Related Work

In this Section, we provide an overview of methods targeting related problems, under various experimental and application settings. Several attempts have been made to utilize machine learning techniques in an automatic system that recognizes different plant species or insects [9], [7], [15], [8], [10], [6], [11], [16]. Existing works, however, focus on a restricted scenario focusing on the classification images depicting a single object. In these settings, it is assumed that a close-range image can be captured that has a single object located at the center of the image patch. For example, the flower classification database [15] contains 17 different flower species with images capturing only few individuals located in the center. Similarly, plant identification through leaf dataset [10] only contains preprocessed patches depicting a single leaf. These datasets were created to focus on the problem of fine-grained classification between similar species, rather than the task of target/object localization within a natural scene that can possibly include multiple object instances at the same time. In addition, the images have been taken with high resolution in an controlled environment, in which the variation in imaging conditions is minimal.

Works in [7], [9] develop automatic fruit detection systems in a harvesting robot. The resemblance between our work and the fruit detection problem is that both deal with natural outdoor imagery that can contain multiple object instances, each of which only accounts for a relatively small area with respect to the whole scene. While flowers and fruits are usually salient objects in an orchard scene, the appearance of flower visitors in some cases even poses a difficulty for human observers to detect, due to relatively small sizes and potential occlusion.

TABLE I
Dataset statistics

#Frames	#Flowers (train/test)	#Visitors (train/test)
650	4590 / 664	441 / 110

III. Flower and Visitor Recognition System

A. Dataset

Data used in our work were collected using time-lapse photography during fieldwork in Greenland near Narsarsuaq during June-July 2017. Greenland is an ideal test location because species diversity of potential flower visitors is remarkably low even compared to other parts of the Arctic. We focus on the widespread artic plant species Dryas integrifolia. At the study site, the most common flower visitors are syrphid flies (Syrphidae, Diptera). Since flower visitors only appear at a specific time during the day, and the appearance time varies from day to day, in order to collect images for the dataset preparation step, a digital camera is setup above a flower bed to capture timelapse video throughout the days. We used two different video sequences captured from cameras placed above two different flower beds, to serve as training and validation data. The use of different validation and training sequences allowed us to measure the generalization of the developed system on unseen data capturing the same flower species in different conditions. From the two video sequences, we sampled frames during a day with varying illuminating and shading conditions to ensure a diversity of outdoor imagery effects in both training and validation sets. The images are recorded with Wingscapes TimelapseCam Pro cameras and stored in JPEG format at 6080×3420 resolution. From the two sets of images coming from two different flower beds, hereby referred to as Set1 and Set2, we annotated the flowers and visitors with rectangular bounding boxes. Example image with annotated bounding boxes are shown in Figure 1 and the statistics of the annotated dataset are given in Table I

ISBN 978-90-827970-1-5 © EURASIP 2018

B. Flower and Visitor Detection Pipeline

While visitors can appear anywhere within the frame, in this study, we are only interested in those visiting the flowers. Thus, our automatic detection system operates two processing steps: image segmentation for the detection of flowers in the scene, and classification of flower patches based on whether they include a visitor or not. Following such a hierarchical process highly reduces the processing time per video frame since, as will be described in the following, flower detection can be conducted at a much lower image resolution, while the task of visitor detection requires high resolution image patches. This is due to that visitors are relatively small and can undergo potential occlusion. The detection pipeline is illustrated in Figure 2.

Fig. 3. Predicted bounding boxes (red) generated by unsupervised image segmentation in proper illumination condition

For flower detection, we followed the sliding window approach, where windows of 64×64 pixels are classified using a CNN trained on a binary flower/no-flower classification problem. The positive training samples are generated by the annotated bounding box information coming from Set1. Since the number of annotated video frames is small, data augmentation was used during training by randomly expanding/shrinking the bounding box within 10% of each dimension before resizing to 64×64. Negative training samples are generated by randomly sampling image patches having less than 30% overlap with ground-truth boxes. Similar data generation steps are applied to Set2 to serve as validation data on patch-based level. To perform the flower localization step in the full-scale video frames, we resized frames to only 10% of the original resolution, i.e., 608×342, then slid the network across the frame with a stride of 10 pixels to generate the prediction. The output of the flower detection step is a map indicating the probability of a pixel to correspond to a bounding box depicting a flower, which is then thresholded as shown in Figure 5.

For visitor detection, we divided the labeled data in Set2 into training and validation sets. The annotation information was used to extract flower patches that contain a visitor as positive samples and those without visitors as negative samples. These patches were used to train another CNN for visitor detection. During the evaluation, based on the flower probability map created by the flower detection network, the binary mask is generated, and flower patches are generated by drawing tight bounding boxes covering the segments. These flower patches are evaluated by the visitor detection network. In this manner, the visitor detection step is made without exhaustive sliding window search. For the success of the visitor detection step, it is necessary that the flower detection network can capture all flower instances appearing in the input frame.

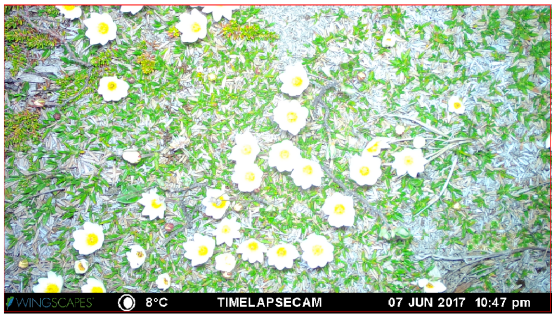

Fig. 4. Predicted bounding box (red box covering entire image) generated by unsupervised image segmentation on bright image

IV. Experiments

A. Experiment protocol

We evaluate our detection system based on two metrics: the misdetection rate (MR) and false positives per image (FPPI). MR is calculated as the ratio between the number of misdetections and the total number of objects in the evaluation set while FPPI is calculated as the ratio between the total number of false positives and the number of input frames. It should be noted that in our flower and visitor detection tasks, both metrics reflect the efficiency of the detection system: an efficient flower detection network should produce low MR to ensure that all flower patches potentially containing a visitor are fed to the visitor detection network, as well as low FPPI to reduce computation overhead of the visitor detection network. Likewise, it is important for the following biological analysis task that all appearances of the visitors are detected, while minimum false positives are returned to minimize manual re-evaluation efforts if needed.

To determine whether the predicted bounding box is a correct detection or misdetection, we used intersection-over-minimum (IOM), which was also used in the task of moth detection [8]. The IOM between two bounding boxes (i.e., the ground-truth and the predicted boxes) is calculated as the ratio between intersection area divided by the minimum area of the two bounding boxes. A detection is counted as correct when IOM > 0.5. While intersection-over-union (IOU) is a popular criterion used in several detection tasks, such as pedestrian detection, we chose to use IOM since we are more interested in the detection and localization of flower and visitor patches within the input frame rather than the precise prediction of the bounding boxes with respect to the ground-truth. For example, if two flowers overlap in the input frame and the ground truth information might contain different bounding boxes for each flower, by boxing the highly probable regions in the flower probability map, our detection system might generate a single

ISBN 978-90-827970-1-5 © EURASIP 2018

bounding box for the two flowers, which may lead to a low IOU score, while being a totally valid and useful prediction in our scenario.

For both flower and visitor detection tasks, we adopted a network architecture similar to the one used in [17], with 12 convolutional layers and a global average pooling layer at the output of the network. Both networks were trained using Adam optimizer [18]. The weight decay regularization was set to 0.0001, and Dropout (with $p = 0.3$) was applied after pooling layers. During training, input patches were randomly shifted as well as horizontally flipped in order to enrich variations of positive and negative sample appearances. We trained our networks for 100 epochs with decaying learning rates $(0.01, 0.005, 0.001, 0.0005, 0.0001)$ after every 20 epochs. The mini-batch size was set to 128 samples.

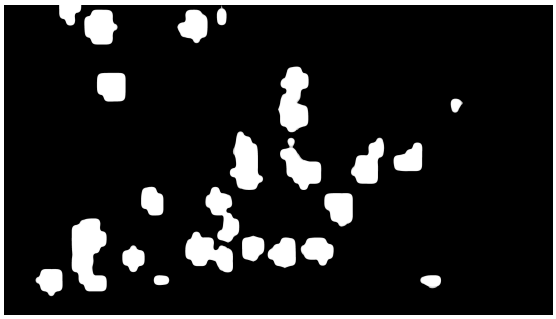

Fig. 5. A flower probability map generated by our system

Fig. 6. Flower patch with occluded visitor

B. Results

Since flowers might have distinctive color information as compared to the background, one might propose to use histogram-based image segmentation techniques that require no annotation information. As we will show in this section, traditional unsupervised image segmentation techniques highly rely on predefined hyper-parameter values. Moreover, varying outdoor imagery conditions greatly affect the segmentation results, forbidding their application in outdoor scenarios. Particularly, we experimented with an unsupervised segmentation applied in HSV color-space capturing the chromaticity information. We followed an approach combining the segmentation masks obtained by thresholding the H and V channels based on thresholds calculated by following [19]. In proper illumination conditions, the unsupervised approach produces reasonable results with few misdetections, as illustrated in Figure 3. On the contrary, the unsupervised approach fails miserably when the brightness is high, as illustrated in Figure 4. The advantage of a supervised system is that invariance in the varying outdoor conditions can be learned to some degrees. This can be seen in Figure 7, where the predicted bounding boxes are generated on the same bright input frame by using the CNN-based detection method.

TABLE II
PERFORMANCE (VALIDATION SET)

	Misdetection Rate (MR)	False Positives Per Image (FPPI)
Flower	8.12%	7.25
Visitor	3.90%	1.41

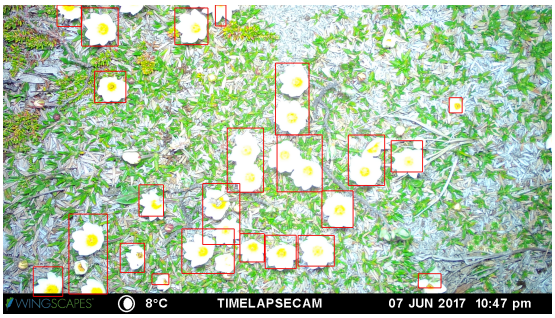

Fig. 7. CNN-generated flower bounding boxes (red) on a bright image

During training, the patch-based performance of both flower and visitor detection networks are relatively high with classification error less than 5%. We are, however, interested in the performance of the system on the input frames, i.e., the full-scale images. Table II shows the performance of both networks on the full-scale images coming from the validation set. It is clear that the evaluated system performs relatively well both in terms of MR and FPPI. Especially the visitor detection network, with only 3.9% misdetections and about 1.5 false positive per image. While the number of false positives returned by the flower detection network is not perfect, this does not affect much the practicality of the system. On the contrary, the FPPI measure in the visitor detection phase plays a more important role since this number reflects the potential overhead imposed on manual re-evaluation, as mentioned in the previous sections. For the visitor detection case, there are intrinsic reasons that can cause an automatic system to return false positives or to fail to detect a visitor. For example, in the process of searching for nectar, a flower visitor might be occluded when observed from the viewpoint of the camera, as illustrated in Figure 6. The inclusion of these cases in the training set could improve the misdetection rate. This, however, might lead the network to falsely classify flowers with dark pistil or stamen as having a visitor when the number of training samples is not high enough.

ISBN 978-90-827970-1-5 © EURASIP 2018

V. Conclusions

In this paper, we described a methodology for flower and visitor detection system using CNNs. Such an approach can be used to reduce the manual efforts required to collect evidence related to biological patterns in such unconstrained environmental settings. As opposed to popular believes that training an efficient CNN requires a huge amount of data, our CNNs, trained on a relatively small number of annotated inputs, showed good performance in terms of misdetection rate and false positives per image. Experiments have shown that the proposed detection pipeline can be used to build an automatic system.

VI. Acknowledgement

This work was supported by a research grant (17523) from VILLUM FONDEN.

References

[1] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, pp. 91–99, 2015.

[2] J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger," *arXiv preprint*, 2016.

[3] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *IEEE International Conference on Computer Vision*, pp. 2980–2988, IEEE, 2017.

[4] M. A. Waris, A. Iosifidis, and M. Gabbouj, "Cnn-based edge filtering for object proposals," *Neurocomputing*, vol. 266, pp. 631–640, 2017.

[5] G. Li, Y. Xie, L. Lin, and Y. Yu, "Instance-level salient object segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 247–256, IEEE, 2017.

[6] J. Raitoharju, E. Riabchenko, K. Meissner, I. Ahmad, A. Iosifidis, M. Gabbouj, and S. Kiranyaz, "Data enrichment in fine-grained classification of aquatic macroinvertebrates," in *International Conference on Pattern Recognition: Workshop on Computer Vision for Analysis of Underwater Imagery*, pp. 43–48, 2016.

[7] S. Bargoti and J. Underwood, "Deep fruit detection in orchards," in *IEEE International Conference on Robotics and Automation*, pp. 3626–3633, IEEE, 2017.

[8] W. Ding and G. Taylor, "Automatic moth detection from trap images for pest management," *Computers and Electronics in Agriculture*, vol. 123, pp. 17–28, 2016.

[9] I. Sa, Z. Ge, F. Dayoub, B. Upcroft, T. Perez, and C. McCool, "Deepfruits: A fruit detection system using deep neural networks," *Sensors*, vol. 16, no. 8, p. 1222, 2016.

[10] S. H. Lee, C. S. Chan, P. Wilkin, and P. Remagnino, "Deep-plant: Plant identification with convolutional neural networks," in *IEEE International Conference on Image Processing*, pp. 452–456, IEEE, 2015.

[11] E. Riabchenko, K. Meissner, I. Ahmad, A. Iosifidis, V. Tirronen, M. Gabbouj, and S. Kiranyaz, "Learned vs. engineered features for fine-grained classification of aquatic macroinvertebrates," in *International Conference on Pattern Recognition*, pp. 2276–2281, 2016.

[12] P. J. CaraDonna, A. M. Iler, and D. W. Inouye, "Shifts in flowering phenology reshape a subalpine plant community," *Proceedings of the National Academy of Sciences*, vol. 111, no. 13, pp. 4916–4921, 2014.

[13] L. Deng, W. Shen, Y. Lin, W. Gao, and J. Lin, "Surveillance camera-based monitoring of plant flowering phenology," in *International Conference on Geo-Informatics in Resource Management and Sustainable Ecosystems*, pp. 273–283, Springer, 2016.

[14] M. J. Couvillon, C. M. Walter, E. M. Blows, T. J. Czaczkes, K. L. Alton, and F. L. Ratnieks, "Busy bees: variation in insect flower-visiting rates across multiple plant species," *Psyche: A Journal of Entomology*, vol. 2015, 2015.

[15] M.-E. Nilsback and A. Zisserman, "A visual vocabulary for flower classification," in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 1447–1454, IEEE, 2006.

[16] J. Arje, S. Karkkainen, K. Meissner, A. Iosifidis, T. Ince, M. Gabbouj, and S. Kiranyaz, "The effect of automated taxa identification errors on biological indices," *Expert Systems with Applications*, vol. 72, pp. 108–120, 2017.

[17] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," *arXiv preprint arXiv:1412.6806*, 2014.

[18] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[19] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

ISBN 978-90-827970-1-5 © EURASIP 2018