Computers and Electronics in Agriculture 123 (2016) 17–28

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Automatic moth detection from trap images for pest management

Weiguang Ding, Graham Taylor *

University of Guelph, Canada

ARTICLE INFO

Article history:
Received 4 October 2015
Received in revised form 10 February 2016
Accepted 11 February 2016
Available online 23 February 2016

Keywords:
Precision agriculture
Integrated pest management
Pest control
Trap images
Object detection
Convolutional neural network

ABSTRACT

Monitoring the number of insect pests is a crucial component in pheromone-based pest management systems. In this paper, we propose an automatic detection pipeline based on deep learning for identifying and counting pests in images taken inside field traps. Applied to a commercial codling moth dataset, our method shows promising performance both qualitatively and quantitatively. Compared to previous attempts at pest detection, our approach uses no pest-specific engineering which enables it to adapt to other species and environments with minimal human effort. It is amenable to implementation on parallel hardware and therefore capable of deployment in settings where real-time performance is required.

1. Introduction

Monitoring is a crucial component in pheromone-based pest control (Carde and Minks, 1995; Witzgall et al., 2010) systems. In widely used trap-based pest monitoring, captured digital images are analysed by human experts for recognizing and counting pests. Manual counting is labour intensive, slow, expensive, and sometimes error-prone, which precludes reaching real-time performance and cost targets. Our goal is to apply state-of-the-art deep learning techniques to pest detection and counting, effectively removing the human from the loop to achieve a completely automated, real-time pest monitoring system.

Plenty of previous work has considered insect classification. The past literature can be grouped along several dimensions, including image acquisition settings, features, and classification algorithms. In terms of image sources, many previous methods have considered insect specimens (Kang et al., 2012, 2014; Arbuckle et al., 2001; Weeks et al., 1999; Tofilski, 2004; Wang et al., 2012). Specimens are usually well preserved and imaged in an ideal lab environment. Thus specimen images are consistent and captured at high resolution. In a less ideal but more practical scenario, some other works attempt to classify insects collected in the wild, but imaged under laboratory conditions (Larios et al., 2008, 2010; Martinez-Munoz et al., 2009; Lytle et al., 2010; Al-Saqer et al., 2010; Cho et al., 2007; Mayo and Watson, 2007). In this case, image quality is usually worse than the specimen case, but researchers still typically have a chance to adjust settings to control image quality, such as imaging all of the insects under a standard orientation or lighting.

From an algorithmic perspective, various types of features have been used for insect classification, including wing structures (Kang et al., 2012, 2014; Arbuckle et al., 2001; Weeks et al., 1999; Tofilski, 2004), colour histogram features (Le-Qing and Zhen, 2010; Kaya and Kayci, 2014), morphometric measurements (Fedor et al., 2008; Yaakob and Jain, 2012; Tofilski, 2004; Wang et al., 2012), local image features (Le-Qing and Zhen, 2010; Kaya and Kayci, 2014; Wen et al., 2009; Wen and Guyer, 2012; Lu et al., 2012; Larios et al., 2011), and global image features (Xiao-Lin et al., 2009). Different classifiers were also used on top of these various feature extraction methods, including support vector machines (SVM) (Wen et al., 2009; Wang et al., 2012; Larios et al., 2010), artificial neural networks (ANN) (Wang et al., 2012; Kaya and Kayci, 2014; Fedor et al., 2008), k-nearest neighbors (KNN) (Xiao-Lin et al., 2009; Wen and Guyer, 2012), and ensemble methods (Larios et al., 2008; Martinez-Munoz et al., 2009; Wen and Guyer, 2012). In general, however, these proposed methods were not tested under real application scenarios, for example, images from real traps deployed for pest monitoring.

Object detection involves also localizing objects in addition to classification. A few attempts have been made with respect to insect detection. One option is to perform a "sliding window" approach, where a classifier scans over patches at different locations of the image. This technique was applied for inspection of bulk wheat samples (Zayas and Flinn, 1998), where local patches from the original image were represented by engineered features and classified by discriminant analysis. Another work on bulk grain

* Corresponding author.

http://dx.doi.org/10.1016/j.compag.2016.02.003
0168-1699/

inspection (Ridgway et al., 2002) employed different customized rule-based algorithms to detect different objects, respectively. The other way of performing detection is to first propose initial detection candidates by performing image segmentations. These candidates are then represented by engineered features and classified (Qing et al., 2012; Yao et al., 2013). All of these insect detection methods are heavily engineered and work only on specific species under specific environments, and are not likely to be directly effective in the pest monitoring setting.

There are two main challenges in detecting pests from trap images. The first challenge is low image quality, due to constraints such as the cost of the imaging sensor, power consumption, and the speed by which images can be transmitted. This makes most of the previous work impractical, that is, those based on high image quality and fine structures. The second challenge comes from inconsistencies which are driven by many factors, including illumination, movement of the trap, movement of the moth, camera out of focus, appearance of other objects (such as leaves), decay or damage to the insect, appearance of non-pest (benign) insects, etc. These make it very hard to design rule-based systems. Therefore, an ideal detection method should be capable and flexible enough to adapt to different varying factors with a minimal amount of additional manual effort other than manually labelled data from a daily pest monitoring program.

Apart from the insect classification/detection community, general visual object category recognition and detection has been a mainstay of computer vision for a long time. Various methods and datasets (Zhang et al., 2013; Andreopoulos and Tsotsos, 2013) have been proposed in the last several decades to push this area forward. Recently, convolutional neural networks (ConvNets) (LeCun et al., 1998; Krizhevsky et al., 2012) and their variants have emerged as the most effective method for object recognition and detection, by achieving state-of-the-art performance on many well recognized datasets (Ciresan et al., 2012; Lee et al., 2014; Swersky et al., 2013), and winning different object recognition challenges (Russakovsky et al., 2015; Krizhevsky et al., 2012; Szegedy et al., 2015).

Inspired by this line of research, we adopt the popular sliding window detection pipeline with convolutional neural networks as the image classifier. First, raw images are preprocessed with colour correction. Then, trained ConvNets are applied to densely sampled image patches to predict each patch's likelihood of containing pests. Patches are then filtered by non-maximum suppression, after which only those with probabilities higher than their neighbours are preserved. Finally, the remaining patches are thresholded. Patches whose probability meet the threshold are considered as proposed detections.

This paper makes two main contributions. First, we develop a ConvNet-based pest detection method, that is accurate, fast, easily extendable to other pest species, and requires minimal preprocessing of data. Second, we propose an evaluation metric for pest detection borrowing ideas from the pedestrian detection literature.

2. Data collection

In this section, we describe the collection, curation, and preprocessing of images. Details of detection performed on processed images are provided in Section 3.

2.1. Data acquisition

RGB colour images are captured by pheromone traps installed at multiple locations by a commercial provider of pheromone-based pest control solutions, whose name is withheld by request. The trap contains a pheromone lure, an adhesive liner, a digital camera and a radio transmitter. The pheromone attracts the pest of interest into the trap where they become stuck to the adhesive surface. The digital images are stored in JPEG format at 640×480 resolution, and transmitted to a remote server at fixed time point daily. Codling moths are identified and labelled with bounding boxes by technicians trained in entomology. Only one image from each temporal sequence is labelled and used in this study, so labelled images do not have temporal correlation with each other. As a result, all of the labelled moths are unique. Fig. 1a shows a trap image with all the codling moth labelled with blue bounding boxes. Fig. 1b shows an image containing no moths but cluttered with other types of insects. High resolution individual image patches are shown later in Fig. 11, with their characteristics analysed in Section 5.3.

2.2. Dataset construction

The set of collected images is split randomly into 3 sets: the training set, the validation set and the test set. After splitting, the statistics of each set is roughly the same as the entire dataset, including the ratio between the number of images with or without moths, and number of moths per image. Table 1 provides specific statistics on the entire dataset and the three splits subsequently constructed.

2.3. Preprocessing

Trap images were collected in real production environments, which leads to different imaging conditions at different points in time. This is most apparent in illumination, which can be seen in Fig. 2a. To eliminate the potential negative effects of illumination variability on detection performance, we perform colour correction using one variant (Nikitenko et al., 2008) of the "grey-world" method. This algorithm assumes that the average value of red (R), green (G) and blue (B) channels should equal to each other. Specifically, for each image, we set the gain of the R and B channels as follows:

$$G_{red} = \mu_{red}/\mu_{green}, G_{blue} = \mu_{blue}/\mu_{green} \tag{1}$$

where μ_{red}, μ_{green} and μ_{blue} are the original average intensities of the red, green and blue channels, respectively. G_{red} and G_{blue} are multiplicative gains applied to the pixel intensity values of the red and blue channels, respectively. Fig. 2b shows images processed by the grey-world algorithm. We see that the images are white-balanced to have similar illumination, but still maintain rich colour information which can be a useful cue for detection downstream. In this paper, all images are white-balanced prior to detection.

3. Detection pipeline

The automatic detection pipeline involves several steps, as shown in Fig. 3. We take a sliding window approach, where a trained image classifier is applied to local windows at different locations of the entire image. The classifier's output is a single scalar $p \in [0, 1]$, which represents the probability that a particular patch contains a codling moth. These patches are regularly and densely arranged over the image, and thus largely overlapping. Therefore, we perform non-maximum suppression (NMS) to retain only the windows whose respective probability is locally maximal. The remaining boxes are then thresholded, such that only patches over a certain probability are kept. The location of these patches with their respective probabilities (confidence scores) are the final outputs of the detection pipeline. We now discuss each of these stages in more detail.