Computers and Electronics in Agriculture 152 (2018) 233–241

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

Multi-level learning features for automatic classification of field crop pests☆

Chengjun Xie[a], Rujing Wang[a], Jie Zhang[a], Peng Chen[a,b,*], Wei Dong[c,*], Rui Li[a], Tianjiao Chen[a], Hongbo Chen[a]

[a] Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei 230031, China
[b] Institute of Physical Science and Information Technology, Anhui University, Hefei 230601, China
[c] Agricultural Economy and Information Research Institute, Anhui Academy of Agricultural Sciences, Hefei 230031, China

ARTICLE INFO

Keywords:
Pest classification
Unsupervised feature learning
Dictionary learning
Feature encoding

ABSTRACT

The classification of pest species in field crops, such as corn, soybeans, wheat, and canola, is still challenging because of the tiny appearance differences among pest species. In all cases, the appearances of pest species in different poses, scales or rotations make the classification more difficult. Currently, most of the classification methods relied on hand-crafted features, such as the scale-invariant feature transform (SIFT) and the histogram of oriented gradients (HOG). In this work, the features of pest images are learned from a large amount of unlabeled image patches using unsupervised feature learning methods, while the features of the image patches are obtained by the alignment-pooling of low-level features (sparse coding), which are encoded based on a predefined dictionary. To address the misalignment issue of patch-level features, the filters in multiple scales are utilized by being coupled with several pooling granularities. The filtered patch-level features are then embedded into a multi-level classification framework. The experimental results on 40 common pest species in field crops showed that our classification model with the multi-level learning features outperforms the state-of-the-art methods of pest classification. Furthermore, some models of dictionary learning are evaluated in the proposed classification framework of pest species, and the impact of dictionary sizes and patch sizes are also discussed in the work.

1. Introduction

Currently, the manual categorization and identification of pest species by expert entomologists are facing great challenges due to the vast number of pest species in the world. It is partly because the pest identification task is time-consuming and requires expert knowledge of field crops. The thorough understanding of pest species requires the terminology of insect taxonomy and morphological characteristics. Therefore, it is difficult to discriminate pest categories at the species level, which leads to the increase of crop losses or the misuse/overuse of pesticides. Along with the development of computer vision and pattern recognition techniques, automated pest classification has attracted a great deal of attention in recent years and has been widely used in many fields, such as agricultural engineering (Zhao et al., 2012), entomological science (Weeks et al., 1999), and environmental science (Larios et al., 2008). However, it is a challenging task since pest species exhibit large variations. Conventional pest classification methods (Weeks et al., 1999; Russell et al., 2005; Arbuckle et al., 2001; Wen et al., 2009) that were developed with shallow learning (e.g., Support Vector Machines, PCA, Boosting, and Logistic Regression) usually worked well only for the cases of good pest images, such as in uniform illumination, consistent scales or positions of the pests in the images, and similar poses or rotations of the pests. Here, we are focusing on the automatic classification of field crop pests whose images were collected from actual field circumstances, which requires that the identification algorithms be highly robust to various challenges of pest appearances, such as backgrounds, illumination changes, scale and pose changes. The cases of the challenges can be found in supplementary file 1.

Many methods on pest appearance modeling were proposed to address the challenges, including concatenated features of local appearance modeling (Larios et al., 2008; Wen and Guyer, 2012; Wang et al., 2012), scale invariant feature modeling (Solis-Sánchez et al., 2011), shape features using quality threshold ARTMAP modeling (Yaakob and Jain, 2012) and, most recently, sparse representation modeling (Xie et al., 2015). Yalcin (2015) tried to discriminate and classify the insects in the pheromone traps under challenging illumination and environmental conditions, with features extracted by the use of Hu moments

☆ Availability: http://www2.ahu.edu.cn/pchen/web/DLFautoinsects.htm.

* Corresponding authors at: Institute of Physical Science and Information Technology, Anhui University, Hefei 230601, China (P. Chen).
E-mail addresses: cjxie@iim.ac.cn (C. Xie), pchen@ahu.edu.cn (P. Chen), dw06@163.com (W. Dong).

https://doi.org/10.1016/j.compag.2018.07.014
Received 12 October 2016; Received in revised form 4 May 2018; Accepted 7 July 2018
0168-1699/

(Hu), Elliptic Fourier Descriptors (EFD), Radial Distance Functions (RDF) and Local Binary Patterns (LBP). Most of the approaches for pest appearance modeling typically rely on either raw image patches or hand-designed image features (e.g., SIFT features, Lowe, 1999). Since raw pixels or image patches are sensitive to noise and background clutter for a natural image, it is hard to cope with the challenge of appearance variations. To model complex real-world pest appearance for automatic classification, robust and distinctive feature descriptors are required to capture the relevant information of pest appearances. Hand-crafted features, such as SIFT and HOG, have made drastic progresses in many vision tasks, such as object recognition and image matching. They are considered as one milestone in computer vision since they have passed the test of time for good performance. Although hand-designed features are effective for capturing low-level image features, it is difficult to find the appropriate representations of mid and high-level features, such as object parts, which are essentially important for representing images. Moreover, the hand-designed features are also criticized for weaknesses, such as large computational burdens and being incapable of properly accommodating appearance variations among pest species. So far, many feature descriptors have been developed for different data sets and tasks, which lack the generalization ability to cope with the appearance variations in different applications. In addition, the hand-crafted features typically require the domain knowledge of the related images for different application scenarios.

Recently, some researchers have focused on multi-level learning features that were extracted from a large amount of unlabeled images by the use of unsupervised machine learning. The results have shown that unsupervised feature learning models outperformed hand-crafted feature representations in many artificial intelligence domains, such as visual recognition (Bo et al., 2013; Le, 2013), natural language processing (Mikolov et al. 2013) and many more. Typical unsupervised feature learning or deep learning approaches can be divided into four categories: sparse coding (Wright et al., 2010), convolutional neural networks (Schmidhuber, 2015), restricted Boltzmann machines (Hinton, 2010), and autoencoders (Zhou et al., 2012). This work mainly aims to design a robust feature learning model that confronts the aforementioned challenges by the use of sparse coding. Particularly motivated by the successes of the works such as Coates et al. (2011), Coates and Ng (2011) and Bo et al. (2013), this paper adopts the deeply learned features into a multi-level classification framework for the automatic classification of field crop pests.

The outline of our model is illustrated in Fig. 1. First, a dictionary is trained from a large amount of unlabeled image patches using unsupervised feature learning methods. Second, the low-level features (namely, sparse coding) are computed from many labeled pest image patches by the learned dictionary. Third, the low-level features are then spatially alignment-pooled to form patch-level features using a multi-level operating strategy. Finally, a multi-level classification framework is constructed by learning the multiple patch-level features of the labeled samples for pest categorization and recognition.

This work is closely related to that of Xie et al. (2015). There are major differences between our work and Xie et al. (2015), although they both used unsupervised feature learning to obtain the learning dictionary. On the one hand, instead of using raw features (e.g., colors, shapes, and textures) as image descriptors, our proposed deep features are learned from a large scale of small image patches, which are randomly extracted from natural images. Moreover, this work applies multiple levels of pest image representations to identify pest species. On the other hand, Xie et al. (2015) considered the sparse-coding histograms of pest species as their features and ignored the spatial structural information in pest images.

The main contributions of this paper are as follows:

- a highly discriminative and robust pest object representation with multi-level learning features,
- a multi-level classification framework with alignment-pooled features using a multi-level operating strategy, and
- a large pest dataset of 40 categories with high quality that were labeled by agricultural experts.

2. Materials and methods

2.1. Dataset collection

We collected approximately 4500 pest images covering most of the species found in several common field crops, including corn, soybean, wheat, and canola. Most of these pest images were captured under real conditions in several experimental fields of the Anhui Academy of Agricultural Sciences in China. Our pest dataset (D0) contains 40 different pest species. The details of the dataset are listed in Table 1, and some typical images are shown in Fig. 5. All images were captured by the use of digital cameras (e.g., Canon, Nikon, and mobile devices). To eliminate the potential negative effects of illumination variability, all sample images were preprocessed with uniform illumination settings in crop field situations, as was done in Wen and Guyer (2012). Furthermore, they were normalized and rescaled to 200 * 200 pixels in this study for computational efficiency. The set of collected images for each insect species was split randomly into 2 subsets: the training set (with approximately 30 images) and the test set (with the remaining pest images). The pest dataset in this paper can be found at our website (http://www2.ahu.edu.cn/pchen/web/DLFautoinsects.htm).

2.2. Pest image representations

The pest image representations in this work with multi-level learning features consist of the following three steps: (i) The unsupervised dictionary learning step, (ii) The feature encoding step, and (iii) The multi-level sampling step.

2.2.1. Unsupervised dictionary learning

Given a set of natural pest images $Y = [y_1, y_2, ..., y_m]$, N overlapped local image patches $x_i \in R^n$, $(i = 1, 2, ..., N)$ are extracted with a spatial layout from the pest images. To make the local patches bright and normalized, each patch is normalized by subtracting the mean of Y and dividing the standard deviation of its elements. More details of the preprocessing for local image patches can be found in Coates et al. (2011). Then, the normalized local patches are used as training data to learn a dictionary using an unsupervised learning algorithm.

Like the KSVD (Aharon et al., 2006), the dictionary $D = [d_1, d_2, ..., d_M] \in R^{n\times M}$ and the associated sparse codes $A = [\hat{a}_1, \hat{a}_2, ..., \hat{a}_N] \in R^{M\times N}$ in this work are learned from the overlapped local image patches $X = [x_1, x_2, ..., x_N] \in R^{n\times N}$ by solving the following optimization problem:

$$\min_{D,A} \sum_{i}^{N} \|x_i - D\hat{a}_i\|_F^2, \ s.\ t.\ \ \|\hat{a}_i\|_0 \leqslant k, \forall\ i, \tag{1}$$

where $\|\cdot\|_F$ denotes the Frobenius norm, $\|\cdot\|_0$ is the l_0 norm that counts non-zero elements in the sparse codes $\hat{a}_i$, k is the sparsity level that controls the number of the non-zero elements, n is the dimension of the image patch vector, M is the dictionary size, and each column d_j in D is a codeword. Therefore, the learning dictionary D is obtained to encode the features of pest images using Eq. (1).

2.2.2. Feature encoding

For a given pest image y_i, the image is partitioned into overlapping square patches $[x_1, x_2, ..., x_K] \in R^{n\times K}$ using a uniform grid. Given the learning dictionary D, the problem of feature encoding is to find the sparse code $\hat{a}_i$ of each patch x_i. This leads to solving the following optimization problem.