Computers and Electronics in Agriculture 119 (2015) 123–132

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

Automatic classification for field crop insects via multiple-task sparse representation and multiple-kernel learning ☆

Chengjun Xie [b,1], Jie Zhang [b,*], Rui Li [b], Jinyan Li [c], Peilin Hong [b], Junfeng Xia [a,*], Peng Chen [a,*]

[a] Institute of Health Sciences, Anhui University, Hefei 230601, China
[b] Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei 230031, China
[c] Advanced Analytics Institute, University of Technology, Sydney, New South Wales, Australia

A R T I C L E I N F O

Article history:
Received 27 January 2015
Received in revised form 21 October 2015
Accepted 22 October 2015

Keywords:
Insect classification
Multiple-task sparse representation of insect objects
Multiple-kernel learning
Sparse coding

A B S T R A C T

Classification of insect species of field crops such as corn, soybeans, wheat, and canola is more difficult than the generic object classification because of high appearance similarity among insect species. To improve the classification accuracy, we develop an insect recognition system using advanced multiple-task sparse representation and multiple-kernel learning (MKL) techniques. As different features of insect images contribute differently to the classification of insect species, the multiple-task sparse representation technique can combine multiple features of insect species to enhance the recognition performance. Instead of using hand-crafted descriptors, our idea of sparse-coding histograms is adopted to represent insect images so that raw features (e.g., color, shape, and texture) can be well quantified. Furthermore, the MKL method is proposed to fuse multiple features effectively. The proposed learning model can be optimized efficiently by jointly optimizing the kernel weights. Experimental results on 24 common pest species of field crops show that our proposed method performs well on the classification of insect species, and outperforms the state-of-the-art methods of the generic insect categorization.

1. Introduction

There are over a million species of insects in the world. Manual categorization and identification of these species is time-consuming and requires expert knowledge of field crops. Traditionally, insect categorization has mainly relied on manual identification by expert entomologists. However, for laymen without a thorough understanding of the terminology of insect taxonomy and morphological characteristics, it is hard to discriminate insect categories at the species level. Therefore, effective identification of insects is a key issue that needs to be well addressed. Computer vision techniques play a crucial role in many research fields such as entomological science (Weeks et al., 1999), environmental science (Larios et al., 2008), and agricultural engineering (Zhao et al., 2012). In this case, computer vision methods could be a feasible way of solving the problem of automated insect categorization and identification. Although many insect categorization approaches have been proposed and have shown to be successful under various scenarios, insect identification is challenging because the variability of colors, textures, and shapes within a single species is very large relative to the variability between species.

There is a rich literature on image or insect appearance modeling (Larios et al., 2008; Luis et al., 2011; Yaakob and Jain, 2012). See an example in Fig. 1. Color histogram is perhaps the simplest way to represent object appearance in the classification of insect species. However, it misses the spatial information of object appearance, making the method sensitive to noise as well as appearance variations in insect categorization. It is widely understood that instead of using a single feature from insect species, combining complementary features such as color, shape, and texture information should be more effective to discriminate among various insect species. An issue is that the performance of feature-based fusion methods, which depend mainly on simple feature extraction and fusion, may deteriorate after the reduction of data dimensionality. In this paper, we propose a robust insect-categorization model that confronts the aforementioned difficulties. The novel idea is to use a sparse-coding technique, which creates global feature descriptors for insects, in combination with a multiple-kernel learning (MKL) technique. The work flow of our method can be decomposed into two stages. The first stage focuses on image or insect object representation. At this stage, global color, texture, and shape features of insect images are extracted using the sparse coding technique. The second stage, which deals

☆ Availability: http://www2.ahu.edu.cn/pchen/web/insectRecognition.htm.
* Corresponding authors. Tel.: +86 551 65591467.
E-mail addresses: cjxie@iim.ac.cn (C. Xie), bigeagle@mail.ustc.edu.cn (P. Chen).
[1] Tel.: +86 551 65591467; fax: +86 551 65592420.

http://dx.doi.org/10.1016/j.compag.2015.10.015
0168-1699/

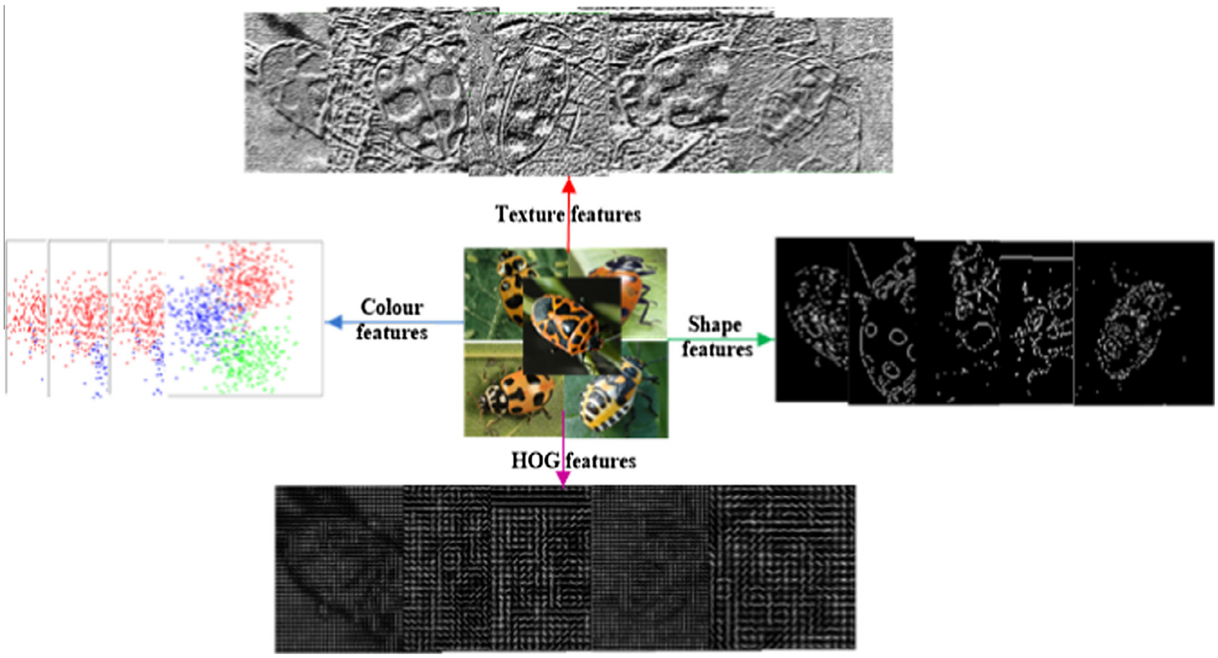

Fig. 1. Visual representations of insect appearance with color, texture, shape, and histogram of oriented gradients (HOG) features.

with effective fusion of multiple insect-categorization features, constructs a kernel-level fusion classifier using all the features.

A novel multiple-task sparse representation of insect objects is proposed by this work (Fig. 2), motivated by the considerable progresses made recently in the research area of sparse representation and coding. For the object representation, multiple over-complete dictionaries with multiple feature modalities of labeled insect images is learned first. Then, multiple modalities of local features are extracted from an insect image, and then the local image patches of the insect object are represented by their sparse codes with the corresponding training dictionary. Despite the fact that insect appearance is modeled using local patches, the global structure information is necessary for accurate insect identification. Finally, insect appearance is represented by concatenating the sparse-coding histograms of all the image patches.

At the second stage, a kernel-level fusion approach with MKL is exploited to classify insects (Fig. 3). In many real classification systems for insect species, a single type of feature is too weak to represent an insect because many features are common to different classes with similar colors or shapes, which leads to ambiguousness in insect classification. To ensure greater discriminative ability, the MKL approach is adopted to combine multiple features via the sparse-coding histograms. Given a set of positive and negative insect samples, multiple modalities of local features are extracted, and then, local image patches of the samples are represented by their sparse codes using the corresponding training dictionary. Finally, an MKL classifier is constructed by learning the sparse-coding histograms of the negative and positive samples for insect categorization and recognition. Compared with existing algorithms for automatic classification of insect species, our technical novelties are as follows:

- the highly discriminative and robust insect object representation with sparse-coding histograms, and
- the combination of multiple complementary features with MKL, where MKL is a tool that represents each image by the use of multiple sets of features in object recognition.

2. Related works

Automated insect identification has been intensively studied over the past two decades, including computer vision-based systems for the classification of insect species (Weeks et al., 1999; O'Neill, 2000; Steinhage et al., 2001; Arbuckle et al., 2001; Wen and Guyer, 2012; Yaakob and Jain, 2012). Weeks et al. (1999) established the digital automated identification system (DAISY) to classify wasp insect images using principal component analysis. To improve classification accuracy, O'Neill (2000) applied DAISY to recognize insect images by analyzing their wing patterns and shapes. Steinhage et al. (2001) developed the automated bee identification system (ABIS) using linear discriminate analysis (LDA) technique. Instead of using LDA, Arbuckle et al. (2001) proposed an improved ABIS system using support vector machine (SVM) and kernel discriminate analysis based on geometric features of wings (such as length, angle, and area). Moreover, many literature works have focused on constructing object appearance models, a key part of object classification. Generally, based on their appearance models, most object feature descriptors can be categorized as either global features or local features. Russell et al. (2005) adopted global features (including color, texture, and geometry) to classify insect images and obtained good results using high-quality images. However, because the features are very sensitive to rotation, scale, translation, and viewpoint changes, this classification method did not work well on objects with large intra-species variation or high inter-species similarity. To address these issues, Wen et al. (2009a) developed a local feature-based insect identification scheme to account for variations in insect appearance. Furthermore, Wen and Guyer (2012) devised an image-based automated insect identification and classification method using three models: an invariant local feature model, a global feature model, and a hierarchical combination model. Luis et al. (2011) extended the LOSS algorithm (Solis-Sánchez et al., 2009) for analyzing the geometrical characteristics of insects to improve insect classification. Wang et al. (2012) adopted artificial neural

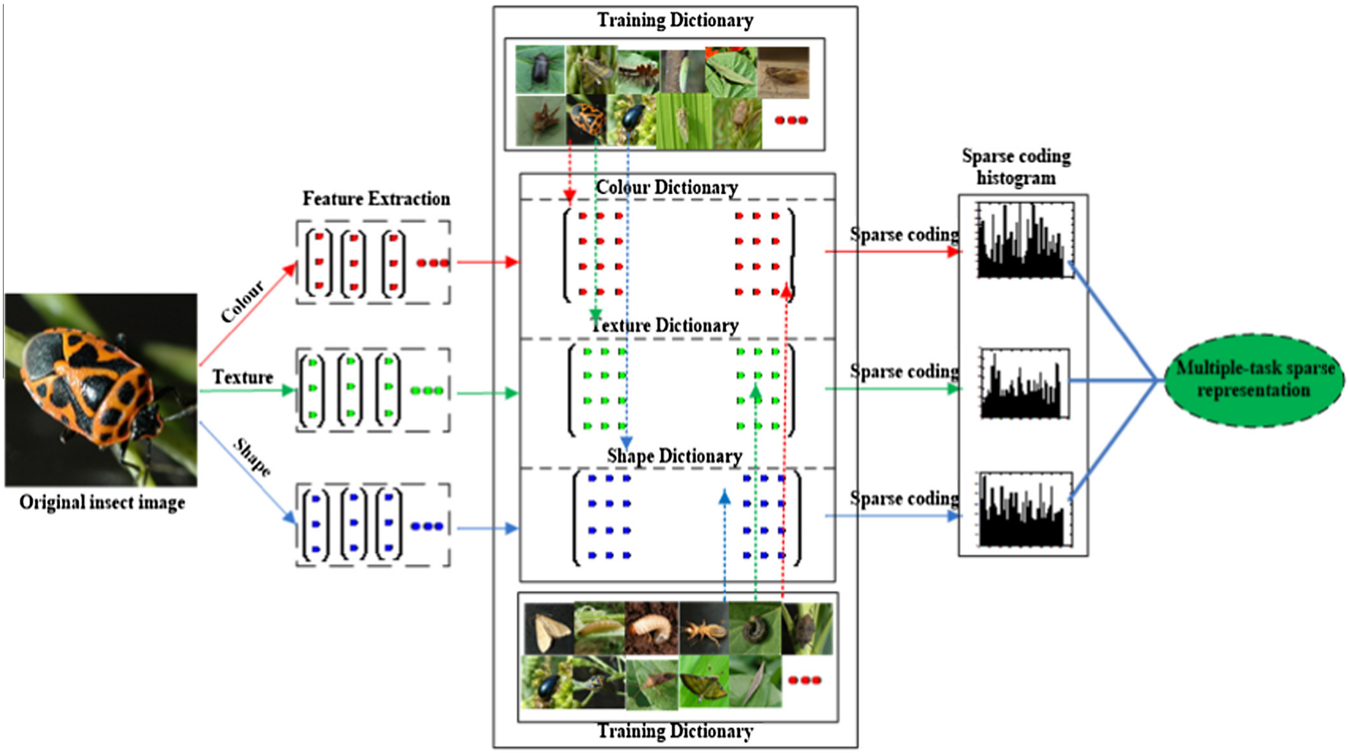

Fig. 2. Insect object representation of our multiple-task sparse representation approach. Given an insect image, multiple feature modalities are extracted firstly. Each feature is then represented as a linear combination of the corresponding training feature dictionary. Then, a multiple-task sparse representation with sparse-coding histogram is used to represent insect image in a joint sparse way over all the features.

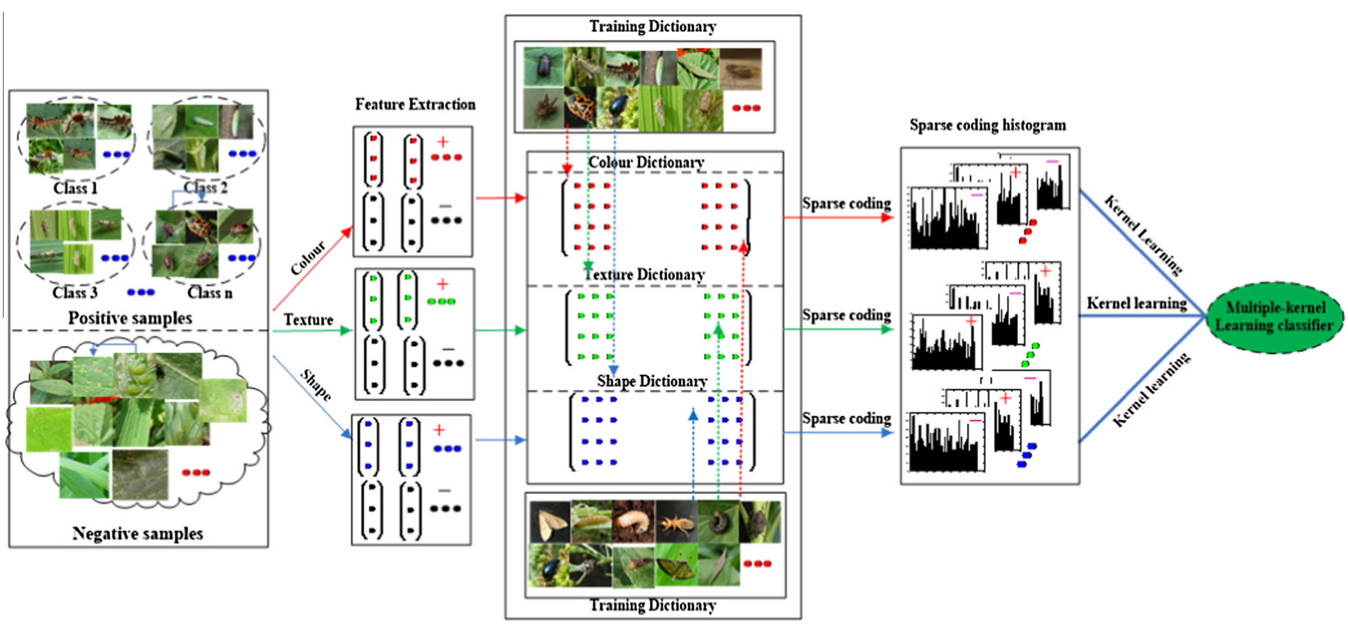

Fig. 3. Kernel-level feature fusion with MKL. Given positive and negative insect samples, multiple modalities of features are extracted. Local image patches of the samples are then represented by their sparse codes using the corresponding training dictionary. Finally, an MKL classifier is constructed by learning the sparse-coding histograms of the negative and positive samples for insect categorization and recognition.

networks (ANNs) and SVM as pattern-recognition methods and designed a new automatic identification system for insect specimen images at the order level. Moreover, Yaakob and Jain (2012) investigated the use of the moment-invariant technique in combination with neural networks to classify insect images. Although the aforementioned insect identification systems have had great success to some extent, most of them focused on high-quality images with uniform illumination, consistent position, and top view pose. Another drawback is the limited expressiveness of these models because insect appearances can only be represented by hand-crafted descriptors (such as scale-invariant feature transform (SIFT), HOG, Gabor, and local binary pattern), which make it difficult to handle significant insect view and pose changes simultaneously.

It is well known that natural images can be sparsely represented (Olshausen and Field, 1997) using a sparse linear combination of a few elements from a trained dictionary. In contrast to most existing insect-classification methods that directly operate on low-level features or cues, sparse coding can learn insect appearances from raw features to quantify insect appearances by means of sparsity (Coates and Ng, 2011). Sparse coding on top of raw patches or features has been applied successfully to face recognition (Yang et al., 2011; Zhang et al., 2011) and generic object recognition (Yang et al., 2009a,b; Zhang et al., 2011; Wang et al., 2010; Ge et al., 2014) to achieve state-of the-art performance. These sparse-based appearance models originate from a pioneering work on face recognition (Wright et al., 2009), which introduced the discriminative nature of sparse representation to a classifier for face recognition and achieved higher recognition rates in constrained experiments. However, Wright's method cannot handle appearance variations and misalignment errors owing to its use of a single feature type, which usually fails to recognize true objects from complex backgrounds.

Recent studies showed that combining multiple types of features can improve recognition performance in visual classification (Yuan et al., 2012). However, direct application of multiple types of features to a classifier for object recognition may not yield the required results because the underlying assumption behind many multi-feature learning algorithms is that those features are highly related, and that the related features can deteriorate classifier performance in insect classification. Thus, a key issue is how to fuse those multiple features. One popular method used in machine learning is MKL, originally proposed by Scholkopf and Smola (2001). A number of studies have shown that MKL is a useful tool for object recognition, where each image is represented by multiple sets of features. One merit of MKL is that it can linearly combine similarity functions between different feature sets such that the combined learning model yields an improvement in recognition performance (Yang et al., 2009a,b; He et al., 2014).

This paper proposes a novel robust insect appearance model for insect classification, which is different from other methods in the following two aspects. First, for the insect image representation, our method represents the target appearance with multiple sets of features. The representation is constructed automatically via sparse coding rather than using traditional feature descriptors with hand-crafted formation. The insect image representation created using our method provides a more flexible mechanism and richer expressiveness, which help solve the abovementioned problems. Second, for the issue of multi-feature fusion, sparse coding of insect images is used to construct training data and the MKL method is adopted to learn a classifier for insect image classification. Such an insect-classification model can effectively alleviate manual load as well as disturbances in natural environments and, therefore, can yield more robust results than other the conventional categorization models, which use raw features of insect images only.

3. Materials and methods

This section presents the details of our insect-classification algorithm. The proposed method consists of two major components: (i) feature-level fusion based on joint sparse coding, and (ii) a multiple-kernel classifier using sparse-coding histograms.

3.1. Sparse representation: a brief introduction

Sparse representation has attracted a great deal of attention in recent years and has been used widely in many fields such as visual classification (Zhang et al., 2011; Wang et al., 2010; Ge et al., 2014). Given a testing image feature $y \in R^n$, it can be represented as a linear representation of basic elements from a dictionary $D \in R^{n\times c}$, which is a stack of n columns of training image feature vectors with dimension c. A representation of the testing image feature y based on dictionary D is any vector $x \in R^c$ that satisfies

$$y = Dx + z, \tag{1}$$

where the dictionary D is said to be over-complete if $n < c$, and z is a noise term with bounded energy $\|z\|_2 < \varepsilon$. However, the solution of x is generally non-sparse with many nonzero elements. To obtain the sparse coding of the input image feature y, the problem can be described formally as follows:

$$\widehat{x_0} = \arg\min \|x\|_0 \quad \text{subject to} \quad \|y - Dx\|_2 < \varepsilon, \tag{2}$$

where $\|\cdot\|_0$ is the l_0 norm, which counts the number of nonzero elements, $\|\cdot\|_2$ is the l_2 norm, and ε denotes the level of reconstruction error. Because the combinatorial l_0-norm minimization is an NP-hard problem, the l_1-norm minimization is used and the problem is formulated as

$$\widehat{x_1} = \arg\min \|x\|_1 \quad \text{subject to} \quad \|y - Dx\|_2 < \varepsilon, \tag{3}$$

where the solution of $\widehat{x_1}$ is the sparse coding of the input image feature y.

3.2. Multiple-task sparse coding for insect image representation

The sparse representation model described in Section 3.1 was originally developed for single feature-based image representation. In this section, this model is generalized to deal with multiple feature-based sparse coding for insect image representation and it is called multiple-task sparse coding. To obtain the multiple-task insect image representation based on sparse coding, m types of visual features, such as color, shape, HOG, scale-invariant feature transform (SIFT), and texture, are used. The widely used global features are color and shape. They can be used to describe entire images. In this work, color histograms (Swain and Ballard, 1991) and an united moment-invariant technique (Yinan et al., 2003) are adopted to extract the color and shape features of insect images. The other three features HOG, SIFT, and texture, are local and are invariant to scale, rotation, illumination, and partial viewpoint change. HoG is used to represent the local appearance of an object based on the local distribution of gradient orientation (Dalal et al., 2006). SIFT was introduced by Lowe (2004) to describe local regions of interest. It calculates local orientation histograms in a local window, where a 128-element SIFT vector is used to describe the local neighborhood of each point of interest. The last feature of texture is very useful for insect recognition and classification. There are many methods to describe the texture of a region. Local binary pattern (LBP) (Ojala et al., 2002) is one of the most widely used methods to extract texture features, and it is adopted in this work.

A training set is considered in which each insect sample is composed of m different feature modalities (color, shape, HOG, SIFT, and texture in this paper). For each feature modality k, $D^k \in R^{n_k\times c}$ denotes the training feature dictionary of the *k-th* feature modality and y^k is the input insect image feature. Inspired by Ge et al. (2014), y^k can be sparsely represented by a linear combination of dictionary D^k for each visual feature, that is, $\hat{x}_1^k = \arg\min \|x^k\|_1$ subject to $\|y^k - D^k x^k\|_2 < \varepsilon,\ k = 1,\ldots,m,$ where $\hat{x}_1^k$ is the sparse coefficient vector (i.e., sparse code) for reconstructing the input insect image feature with visual feature y^k based on dictionary D^k. Finally, these sparse codes $\hat{x}_1^k$ under different feature modalities $k = 1, \ldots, m$ are collected to represent insect image appearance for insect classification.

To make full use of the partial information on an insect image, a local sparse representation is used to model the appearance of insect image patches, where a set of sparse codes is collected to represent the image appearance. Let $P^k = \{p_i^k | i = 1 : T\}$ denote the k-th feature modality of an insect image to be coded, where $p_i^k \in R^n$ is the i-th local image feature extracted from the insect image. In the dictionary D^k, each p_i^k has a corresponding sparse code $\alpha_i \in R^{c\times 1}$, which can be computed as

$$\begin{aligned} \widehat{a}_i^k = \quad & \arg\min \|\alpha_i^k\|_1 \\ & \text{subject to } \|p_i^k - D^k\alpha_i^k\|_2 < \varepsilon, \; k = 1, \ldots, m. \end{aligned} \tag{4}$$

The sparse codes $A = [\widehat{a}_1^k, \widehat{a}_2^k, \ldots, \widehat{a}_T^k]$ of an insect image are computed to represent the insect appearance under different feature modalities. Similar to Liu et al. (2013), sparse-coding histogram is introduced to represent sparse code distribution. The coding histogram is defined under different feature modalities (e.g., color, shape, HOG, SIFT and texture in this paper) as the sum of the non-zero coefficients of basis, as shown below:

$$H^k(q) = C_0 \sum_i^T k\left(\left\|\frac{Y - c_i}{h}\right\|^2\right) |a_{iq}^k|, \tag{5}$$

where $H^k(q)$ is the value of the sparse-coding histogram in the q-th bin for feature modality k, c_i represents local feature i at a certain position in the insect image, Y represents a certain position of the target, $k()$ is an isotropic kernel function that is applied to assign smaller weights to local features far away from the insect image center, C_0 is a normalization constant, and a_{iq}^k is the q-th coefficient of the i-th local feature. The proposed multiple-task sparse representation of insect appearance is shown in Fig. 2.

3.3. MKL for insect classification

To obtain effective feature combinations with the corresponding sparse-coding histograms is a key in class-level insect classification and recognition. In order to combine multiple complementary features for robust classification, a kernel-level feature fusion-based MKL scheme is developed, which can linearly combine similar functions between images.

MKL is a simple yet representative kernel-level fusion method. Traditionally, training a binary classifier that estimates the decision function of SVM (Chen et al., 2007) starts from a training data set with form $\{(x_1, y_1), \ldots, (x_N, y_N)\}$, where $y_i \in \{+1, -1\}$ is a binary label. The decision function of SVM is

$$f(x) = W^T R(x) + b, \tag{6}$$

where $R(x)$ is a mapping of sample x from the input space to a high-dimensional feature space. The classifier can be trained by solving the following quadratic optimization problem:

$$\begin{aligned} & \min_{W,b,\xi} \tfrac{1}{2}\|W\|^2 + C\sum_{i=1}^N \xi_i, \\ & \text{s.t. } y_i(W^T R(x) + b) + \xi_i \geq 1, \quad 0 \leq \text{i} \leq N, \quad \xi_i \geq 0, \end{aligned} \tag{7}$$

where W is the normal vector to the hyperplane, ξ_i is the i-th slack variable, C is the regularization parameter, and b is the bias term. Here parameters (w,b) are determined by solving the following equivalent dual optimization problem:

$$\begin{aligned} & \max\left\{\sum_{i=1}^N \alpha_i - \tfrac{1}{2}\sum_{i,j=1}^N \alpha_i\alpha_j y_i y_j \langle R(x_i), R(x_j)\rangle\right\}, \\ & \text{s.t. } \sum_{i=1}^N y_i\alpha_i = 0, \quad 0 \leq \text{i} \leq N, \quad 0 \leq \alpha_i \leqslant C, \end{aligned} \tag{8}$$

where $O(x_i, x_j) = \langle R(x_i), R(x_j)\rangle$ is a kernel function; Gaussian kernel is the most commonly used function:

$$O(x_i, x_j) = \exp\left(-\delta^2\|x_i - x_j\|^2\right), \tag{9}$$

where δ^2 is the kernel parameter. When a test sample x is input into the classifier, the binary decision function can be expressed as follows:

$$f(x) = \sum_{i=1}^N \alpha_i y_i O(x_i, x) + b. \tag{10}$$

To learn kernel-level combination, MKL is adopted instead of the commonly used single fixed kernel, says Gaussian kernel. The method learns the optimal convex combination of multiple kernels by jointly optimizing kernel weights.

Here, the appearance of an insect image with the corresponding sparse-coding histograms H^k is collected to form the training data under different feature modalities. It is assumed that there are N training samples $\{H_1, \ldots, H_N\}$, where each sample j is described by m types of features, i.e., $H_j = \{H_j^1, \ldots, H_j^m\}$, with the corresponding sparse-coding histograms. The associated class labels for all training samples are represented as an N-dimensional vector y, and each element of this vector is $y_j \in \{+1, -1\}, \; j = 1, \ldots, N$. Then the MKL can be written as follows:

$$\begin{aligned} & \min_{\{W_k\},b,\xi} \tfrac{1}{2}\sum_{k=1}^m \tfrac{1}{\omega_k}\|W_k\|^2 + C\sum_{i=1}^N \xi_i \\ & \text{s.t.} y_i\left(\sum_{k=1}^m \left\langle W_k, R_k(H_i^k)\right\rangle\right) + \xi_i \geqslant 1, 0 \leqslant i \leqslant N, \xi_i \geqslant 0, \\ & \sum_{k=1}^m \omega_k = 1, \;\; \omega_k \geqslant 0 \end{aligned} \tag{11}$$

where ω_k is the weight on the corresponding basic kernel $O_k(H_i^k, H_j^k) = \langle R_k(H_i^k), R_k(H_j^k)\rangle$. Similar to traditional SVM, a test insect sample x can be classified by evaluating the binary decision function of MKL which can be expressed as

$$f(x) = \sum_{k=1}^m \omega_k \sum_{i=1}^N \alpha_i y_i O_k(H_i, x) + b. \tag{12}$$

The kernel-level feature fusion with MKL is shown in Fig. 3.

The flowchart of the proposed insect-classification algorithm with the key components multiple-task sparse representation and MKL is shown in Fig.4.

4. Experiments and discussion

The performance of our proposed classification method is rigorously studied on 24 different types of insects. A comparison of our results with those of the state-of-the-art methods is given at the end of this section.

4.1. Experimental settings

To meet the need for practical insect image identification, we collected insect images covering various species across several common field crops including corn, soybean, wheat, and canola. Samples of 24 common pest species found in field crops were collected, such as *Cifuna locuples*, *Tettigella viridis*, and *Colposcelis signata*. The details of the selected insect species are listed in Table 1, and some cases are shown in Fig. 5. In this work, insect sample images were manipulated by hand. Furthermore, image capture, preprocessing, and region selection also required direct user interaction. All insect images were captured with a color

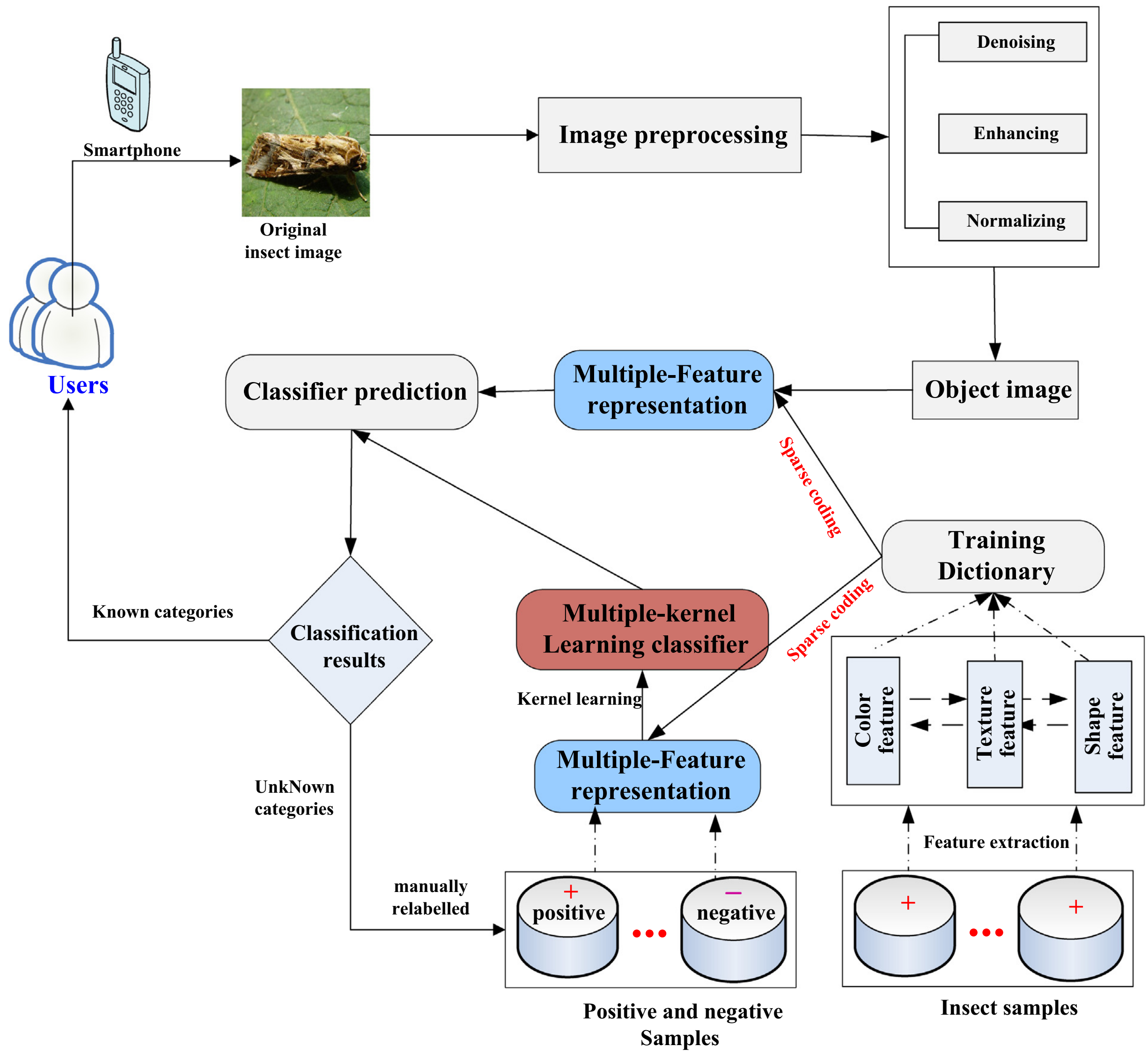

Fig. 4. Flowchart of insect-classification algorithm with multiple-task sparse representation and MKL.

digital camera (e.g., Cannon, Nikon) at a resolution of $1280 * 960$. They were oriented, normalized, and rescaled to $300 * 300$ px in this study for computational efficiency. Furthermore, to achieve good categorization of insect species, the insect samples were preprocessed with uniform illumination settings (detailed information about the settings can be found in Wen and Guyer (2012)) in field situations. The insect dataset in this paper is available at our website.

Three data sets were used in this study for comparison. Our insect dataset contains 24 different types of insects from field-based insect collection, including about 60 images per species. Each species is divided into test and training subsets (with approximately 32 images each), as summarized in Table 1. The second dataset (D1) is from the Butterfly database (Xiao et al., 2012), and the last one (D2) is a large dataset from Wang et al. (2012). For our insect-classification method, the method reported in Nilsback and Zisserman (2008) was employed to extract five different types of features, namely, color, texture, shape, HOG, and SIFT.

Moreover, linear, Gaussian, and polynomial kernels were selected. To determine the appropriate kernel and the parameters for each kernel, a series of experiments was conducted on our dataset. For the Gaussian kernel $O(x_i, x_j) = \exp\left(-\delta^2 \|x_i - x_j\|^2\right), \delta^2$ was set to 10,000 and for the polynomial kernel $O(x_i, x_j) = ((x_i \cdot x_j) + 1)^d$, d was set to 5. The regularization parameter C was chosen from the set $\{0.1, 1, 10, 100, 500, 1000, 1500, 2000\}$. For MKL, C was fixed at 1000 to obtain the best results. Moreover, our experiments were implemented in C++ on a 2.5 GHz machine with 4 GB RAM.

4.2. Parameters analysis

Parameters such as δ^2, d, and the regularization parameter C play important roles in the proposed insect-classification algorithm. In this section, the determination of their values and the corresponding effects on classification performance are described. First, the effects of various δ^2 and d values on the classification performance of insect species were analyzed using different datasets (i.e., our dataset, D1, and D2). To simplify this problem, discrete values of (0.01, 1, 100, 10,000, 15,000, 20,000) and (1, 3, 5, 7, 9, 11) were selected for δ^2 and d, respectively. The experimental results, summarized in Tables 2 and 3, indicate that the method performs better with $\delta^2 = 10,000$ and $d = 5$ than with other parameter values. To determine the optimum value of the regularization

Table 1
Details of 24 analyzed insect species.

Species ID	Insect names	No. of test images	Species ID	Insect names	No. of test images
1	*Cifuna locuples*	25	13	*Eurydema dominulus*	27
2	*Tettigella viridis*	23	14	*Colaphellus bowvingi*	24
3	*Colposcelis signata*	27	15	*Pieris rapae*	26
4	*Maruca testulalis*	25	16	*Eurydema gebleri*	26
5	*Atractomorpha sinensis*	26	17	*Erthesina fullo*	25
6	*Sympiezomias velatus*	28	18	*Chromatomyia horticola*	27
7	*Sogatella furcifera*	25	19	*Eysacoris guttiger*	29
8	*Cletus punctiger*	27	20	*Dolerus tritici*	25
9	*Cnaphalocrocis medinalis*	23	21	*Pentfaleus major*	24
10	*Laodelphax striatellua*	22	22	*Sitobion avenae*	26
11	*Chilo suppressalis*	26	23	*Aelia sibirica*	28
12	*Mythimna separta*	24	24	*Nephotettix bipunctatus*	25

Fig. 5. Sample images for 24 insect species used from field-based insect collection.

Table 2
Effects of δ^2 on categorization accuracy (%) in three datasets employed herein.

Datasets	δ^2					
	0.01	1	100	10,000	15,000	20,000
D1	76.1 ± 1.2	85.3 ± 1.6	90.1 ± 0.7	95.1 ± 1.2	92.0 ± 0.8	90.1 ± 0.3
D2	70.0 ± 1.0	80.0 ± 2.0	85.0 ± 1.1	90.0 ± 1.0	87.2 ± 0.5	85.0 ± 1.4
Our dataset	69.5 ± 0.5	80.5 ± 1.7	84.5 ± 1.6	89.5 ± 1.2	87.4 ± 1.0	85.5 ± 1.0

Table 3
Classification accuracy (%) with different kernel parameters (d) in three datasets employed herein.

Datasets	d				
	1	3	5	7	9
D1	83.2 ± 0.4	87.2 ± 1.0	90.2 ± 1.0	89.2 ± 0.9	87.8 ± 0.9
D2	81.3 ± 1.6	84.3 ± 1.7	87.3 ± 1.4	86.0 ± 0.4	84.3 ± 1.2
Our dataset	80.5 ± 0.5	83.5 ± 0.8	85.5 ± 0.8	84.7 ± 1.1	83.6 ± 0.3

parameter C, MKL was used to evaluate our database with different values of C (0.1, 1, 10, 100, 500, 1000, 1500, 2000). The optimum value of C was determined with respect to categorization accuracy. From Fig. 6, it can be seen that with our dataset, as the value of C increases from 0.1 to 1000, the categorization accuracy of MKL increases; the accuracy decreases for any further increase in the C value. It is suggested that too high a value of C leads to overfitting. This is the case for the other two datasets as well, as in Fig. 6. Hence, the optimum value of C was set to 1000 in this study.

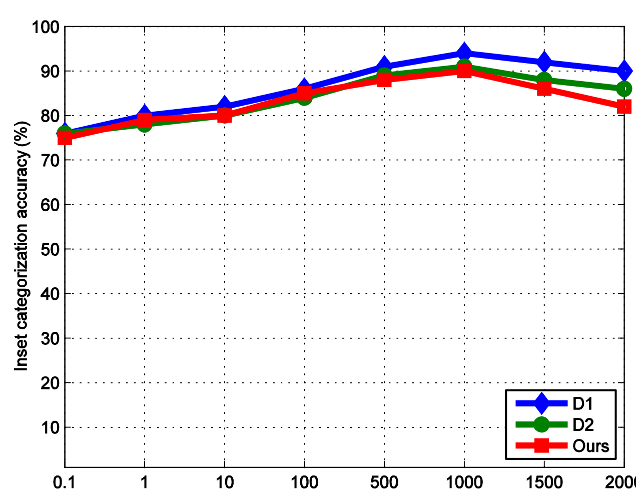

Fig. 6. Categorization accuracy of MKL in three datasets employed herein with different values of C.

Table 4
Categorization accuracy (%) with single feature on insect dataset comprising 24 categories.

Feature	NSC	SVM	SC	LRSC	SCSPM	Proposed
Color	61.3 ± 2.7	60.3 ± 1.9	63.4 ± 2.3	68.3 ± 1.5	65.6 ± 2.0	70.2 ± 1.8
Texture	50.9 ± 2.0	56.8 ± 2.6	59.5 ± 1.9	60.6 ± 0.8	59.3 ± 2.7	63.5 ± 2.0
Shape	68.9 ± 1.3	70.5 ± 3.4	74.3 ± 1.8	78.3 ± 2.8	75.4 ± 3.1	80.2 ± 2.1
SIFT	70.6 ± 2.3	71.8 ± 3.4	77.8 ± 1.5	80.9 ± 1.4	79.5 ± 1.9	81.0 ± 2.4
HOG	52.3 ± 1.7	55.6 ± 2.8	61.1 ± 1.3	62.7 ± 0.9	60.1 ± 3.0	64.1 ± 2.5

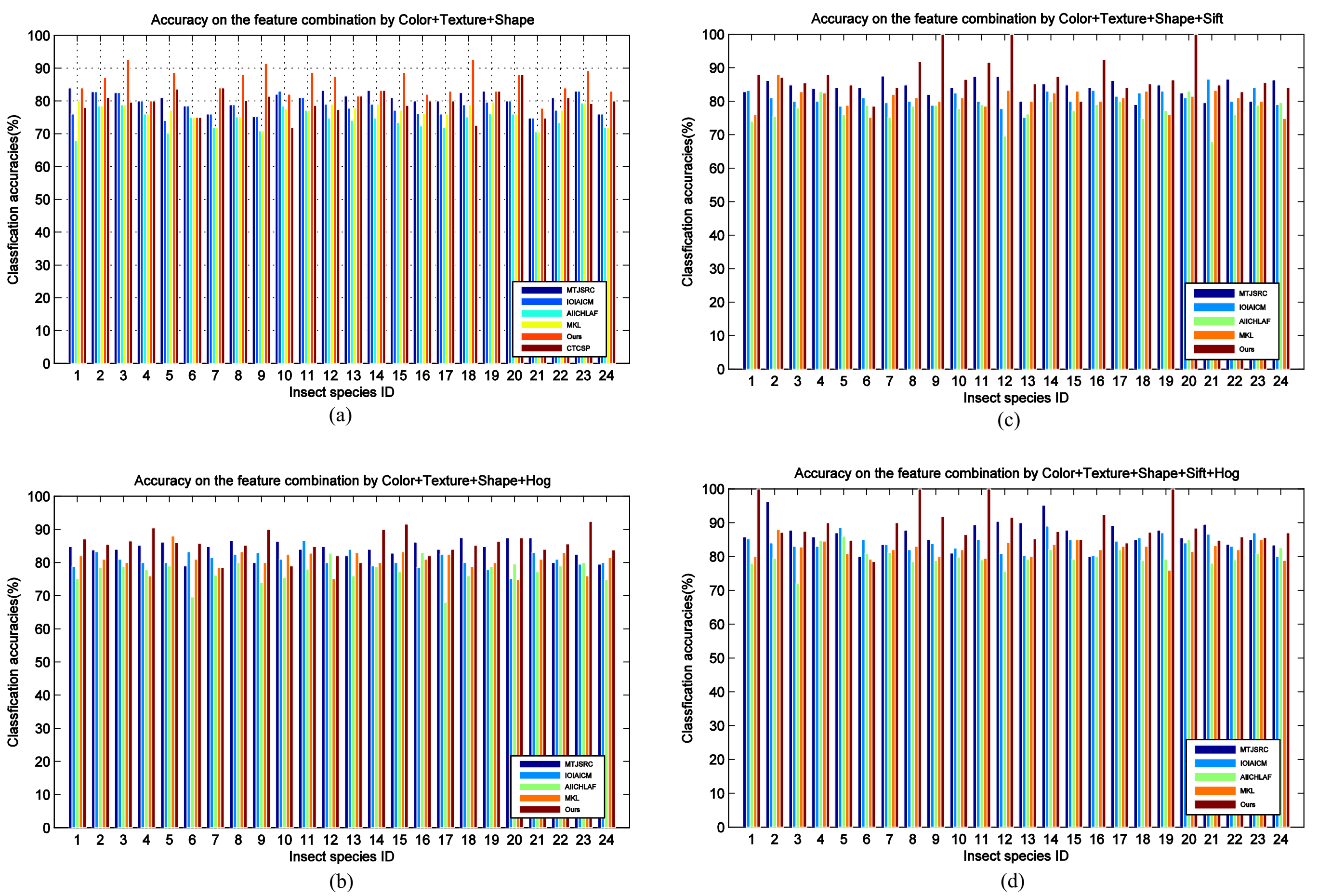

Fig. 7. Accuracies on feature combinations of: (a) color + texture + shape; (b) color + texture + shape + hog; (c) color + texture + shape + sift; (d) color + texture + shape + sift + hog. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.3. Result analysis

We compare our proposed method with a number of recent state-of-the-art image classification methods. The compared methods include SC (Huang et al., 2011), LRSC (Zhang et al., 2013), SCSPM (Yang et al., 2009a,b), MTJSRC (Yuan et al., 2012), IOIAICM (Wen and Guyer, 2012), AIICHLAF (Larios et al., 2008), MKL (Gehler and Nowozin, 2009), as well as two general classifiers, namely, nearest subspace classifier (NSC) and linear SVM.

Table 4 lists the best results of NSC, SVM, SC, LRSC, SCSPM, and the proposed method for each individual feature. Statistically, our method always outperformed the other methods, which use raw features directly (e.g., color, texture, shape, HOG, and SIFT) as training samples. Our method using sparse-coding histograms as training samples achieves higher categorization rates in most cases than SVM and NSC. Additionally, our classification model of insect species performs better than the three recently developed classification methods SC, LRSC, and SCSPM, which use sparse representation directly.

We then evaluated the performance of our method by combining the five features and compared the results with those achieved using the feature combination-based state-of-the-art methods. The categorization accuracy curves with respect to the combinations of the five features and on the 24-insect dataset are shown in Fig. 7. In Fig. 7(a), by combining the color, shape, and texture features, the proposed method achieves higher categorization rates than the other state-of-the-art methods such as CTCSP (Gustavo, 2010), MTJSRC (Yuan et al., 2012), IOIAICM (Wen and Guyer, 2012), AIICHLAF (Larios et al., 2008), and MKL (Gehler and Nowozin, 2009). By using the features of color, texture, shape and SIFT, we can see, in Fig. 7(b)–(c), that all methods obtain higher classification rates than those using combinations of three and four features. Finally, all methods using the combination of all five features performed at their respective best levels, as shown in Fig. 7(d).

Table 5
Categorization accuracy (%) of different methods on two publicly available data sets.

Datasets	NSC	SVM	SC	LRSC	SCSPM	Ours
D1	75.5 ± 2.0	80.1 ± 1.6	85.1 ± 1.7	92.9 ± 1.7	90.8 ± 1.5	97.2 ± 1.0
D2	64.3 ± 1.2	70.0 ± 1.1	79.5 ± 1.7	89.0 ± 1.5	82.3.8 ± 1.3	90.3 ± 1.4

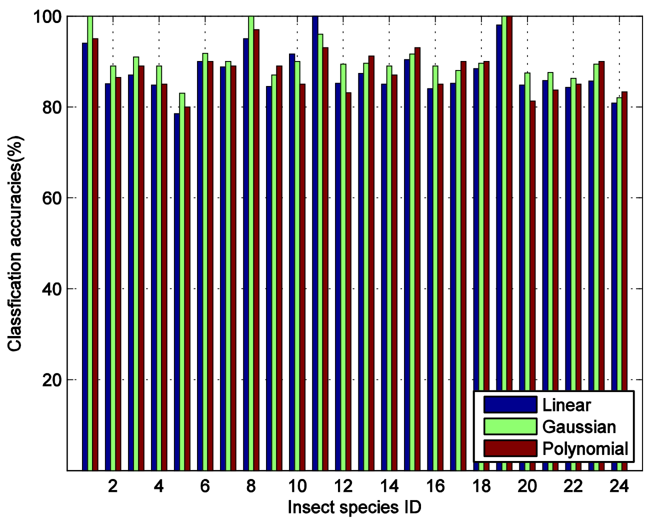

Fig. 8. Recognition accuracy of our approach with different types of kernels on different data sets of insect species.

As expected, compared with the five competing methods, by combining the five features, our method consistently outperforms MTJSRC (Yuan et al., 2012), IOIAICM (Wen and Guyer, 2012), AIICHLAF (Larios et al., 2008), and MKL (Gehler and Nowozin, 2009), which shows the effectiveness of our method in image representation and learning of insect appearance model, i.e., multiple-task sparse representation and MKL.

To further evaluate the insect image description capability of the proposed method, 20 samples from the Butterfly dataset (Xiao et al., 2012) were investigated in this study. The Butterfly dataset was used because it is quite difficult to classify these butterflies based solely on their appearances. The appearance characteristics of butterflies such as shape are very similar across individuals. The main difference between butterflies essentially comes from their wing textures. Furthermore, a larger dataset (Wang et al., 2012) containing 221 insect species from lab-based insect collection was selected for evaluating the recognition accuracy of the proposed method. Table 5 summarizes the categorization accuracies of methods on the two datasets. It can be seen that our method performs better than the other methods on the two publicly available datasets.

Fig. 8 shows the classification accuracy of our method with different types of kernels on different data sets of insect species. It can be seen that the performance of our method is not sensitive to kernel type.

Although our method yields the best performance in the experiments, it takes longer time to train our method, especially to train the dictionary for sparse coding, compared with other feature-combination methods. Thus, a study will develop to enhance the efficiency of our insect-classification algorithm and to investigate the insect appearance representation scheme in sparse coding.

5. Conclusions

This paper proposed a novel method for the classification of insect images. It contains two steps. First, an effective feature description scheme was proposed to represent insect image appearance using sparse-coding histograms with multiple feature modalities. Second, to effectively fuse the features with the corresponding sparse-coding histograms, MKL is adopted to learn the sparse-coding histograms of positive and negative insect images for a multiple-class classifier. The classifier decision function was embedded into an insect-classification system. The proposed method was rigorously evaluated on 24 different types of insect datasets and compared with a number of most recent methods. Experiments on these challenging insect images showed that the proposed system achieves the state-of-the-art performance.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Nos. 31401293, 61300058, 41302261 and 61472282).

References

Arbuckle, T., Schroder, S., Steinhage, V., Wittmann, D., 2001. Biodiversity informatics in action identification and monitoring of bee species using ABIS. In: Proceedings of the 15th International Symposium Informatics for Environmental Protection, pp. 425–430.

Chen, Z., Li, J., Wei, L., 2007. A multiple kernel support vector machine scheme for feature selection and rule extraction from gene expression data of cancer tissue. Artif. Intell. Med. 41 (2), 161–175.

Coates, A., Ng, A., 2011. The importance of encoding versus training with sparse coding and vector quantization. In: Proceedings of the 28th International Conference on Machine Learning (ICML-11), pp. 921–928.

Dalal, N., Triggs, B., Schmid, C., 2006. Human detection using oriented histograms of flow and appearance. In: European Conference on Computer Vision, Lecture Notes in Computer Science, pp. 428–441.

Ge, T., He, K., Sun, J., 2014. Product sparse coding. In: Proceedings of the 2014 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 939–946.

Gehler, P., Nowozin, S., 2009.On feature combination for multiclass object classification. In: IEEE International Conference on Computer Vision, pp. 221–228.

Gustavo E.A.P.A., Batista, Campana, Bilson, Keogh, Eamonn, 2010. Classification of live moths combining texture, color and shape primitives. In 2010 Ninth International Conference on Machine Learning and Applications (ICMLA), pp. 903–906.

He, X., Mourot, G., Maquin, D., Ragot, J., Beauseroy, P., 2014. Multi-task learning with one-class SVM. Neurocomputing 133, 416–426.

Huang, Y., Huang, K., Yu, Y., Tan, T., 2011. Salient coding for image classification. In: Proceedings of the 2011 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1753–1760.

Larios, N., Deng, H., Dietterich, T.G., et al., 2008. Automated insect identification through concatenated histograms of local appearance features. Mach. Vis. Appl. 19 (2), 105–123.

Liu, B., Huang, J., Kulikowski, C., Yang, L., 2013. Robust visual tracking using local sparse appearance model and k-selection. IEEE Trans. Pattern Anal. Mach. Intell. 35 (12), 2968–2981.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision 60, 91–110.

Luis, O.S., Rodrigo, C., Juan, J., Irineo, T., Ramón, G.G., Celina, L.C., Pedro, D.A., 2011. Scale invariant feature approach for insect monitoring. Comput. Electron. Agric. 75, 92–99.

Nilsback, M.-E., Zisserman, A., 2008. Automated flower classification over a large number of classes. In: Proc. IEEE Indian Conf. Computer Vision, Graphics and Image Processing, pp.722–729.

Ojala, T., Pietikäinen, M., Mäenpää, T., 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. 24 (7), 971–987.

Olshausen, B.A., Field, D.J., 1997. Sparse coding with an overcomplete basis set: a strategy employed by V1? Vision. Res. 37 (23), 3311–3325.

O'Neill, M.A., 2000. DAISY: a practical tool for semi-automated species identification. In: MacLeod, N (Ed.), Automated Taxon Identification in Systematics: Theory, Approaches, and Applications. CRC Press/Taylor & Francis Group, Boca Raton/Florida, pp. 101–114.

Russell, K.N., Do, M.T., Platnick, N.I., 2005. Introducing SPIDA-Web: an automated Identification system for biological species. In: Proceedings of Taxonomic Database Working Group Annual Meeting, pp.11–18.
Scholkopf, B., Smola, A.J., 2001. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, Cambridge, MA, USA.
Solis-Sánchez, L.O., García-Escalante, J.J., Casta~neda-Miranda, R., Torres-Pacheco, I., Guevara-González, R.G., 2009. Machine vision algorithm for whiteflies (Bemisia tabaci Genn.) scouting under greenhouse environment. J. Appl. Entomol. 133 (7), 546–552.
Steinhage, V., Arbuckle, T., Schröder, S., Cremers, A.B., Wittmann, D., 2001. ABIS: automated identification of Bee species, BIOLOG workshop. German Programme on Biodiversity and Global Change, Status Report, German Ministry of Education and Research (BMBF), Bonn, pp. 194–195.
Swain, M., Ballard, D., 1991. Color indexing. Int. J. Comput. Vision 7 (1), 11–32.
Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y., 2010. Locality-constrained linear coding for image classification. In: Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 3360–3367.
Wang, J., Lin, C., Ji, L., Liang, A., 2012. A new automatic identification system of insect images at the order level. Knowl.-Based Syst. 33, 102–110.
Weeks, P.J.D., O'Neill, M.A., Gaston, K.J., Gauld, I.D., 1999. Species identification of wasps using principal component associative memories. Image Vis. Comput. 17 (12), 861–866.
Wen, C., Guyer, D., 2012. Image-based orchard insect automated identification and classification method. Comput. Electron. Agric. 89, 110–115.
Wen, C., Guyer, D., Li, W., 2009a. Local feature-based identification and classification for orchard insects. Biosyst. Eng. 104, 299–307.
Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y., 2009. Robust face recognition via sparse representation. IEEE Trans. Pattern Anal. Mach. Intell. 31 (2), 210–227.
Xiao, B., Ma, J.-F., Cui, J.-T., 2012. Combined blur, translation, scale and rotation invariant image recognition by Radon and pseudo-Fourier–Mellin transforms. Pattern Recogn. 45 (1), 314–321.
Yaakob, S.N., Jain, L., 2012. An insect classification analysis based on shape features using quality threshold ARTMAP and moment invariant. Appl. Intell. 37, 12–30.
Yang, J., Yu, K., Gong, Y., Huang, T., 2009. Linear spatial pyramid matching using sparse coding for image classification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1794–1801.
Yang, J., Li, Y., Tian, Y., Duan, L., Gao, W., 2009. Group-sensitive multiple kernel learning for object categorization. In: IEEE 12th International Conference on Computer Vision, pp. 436–443.
Yang, M., Zhang, L., Yang, J., Zhang, D., 2011. Robust sparse coding for face recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 625–632.
Yinan, S., Liu, W., Yuechao, W., 2003. United moment invariants for shape discrimination. In: IEEE Proceedings on International Conference on Robotics, Intelligent Systems and Signal Processing, pp. 88–93.
Yuan, X.-T., Liu, X., Yan, S., 2012. Visual classification with multitask joint sparse representation. IEEE Trans. Image Process. 21 (10), 4349–4360.
Zhang, L., Yang, M., Feng., X.C., 2011. Sparse representation or collaborative representation which helps face recognition? In: IEEE International Conference on Computer Vision, pp. 471–478.
Zhang, T., Ghanem, B., Liu, S., Ahuja, N., 2013. Low-rank sparse coding for image classification. In: IEEE International Conference on Computer Vision, pp. 281–288.
Zhao, Y., He, Y., Xu, X., 2012. A novel algorithm for damage recognition on pest-infested oilseed rape leaves. Comput. Electron. Agric. 89, 41–50.