Insect Species Recognition using Sparse Representation

An Lu Student[1]
alu@nlpr.ia.ac.cn
Xinwen Hou Prof[1]
xwhou@nlpr.ia.ac.cn
Xiaolin Chen Prof[2]
xlchen@ioz.ac.cn
Chenglin Liu Prof[1]
liucl@nlpr.ia.ac.cn

[1] Institute of Automation
Chinese Academy of Sciences
Beijing, China

[2] Institute of Zoology
Chinese Academy of Sciences
Beijing, China

Abstract

Insect species recognition is a typical application of image categorization and object recognition. Unlike generic image categorization datasets (such as the Caltech 101 dataset) that have large variations between categories, the difference of appearance between insect species is so small that only some entomologist experts can distinguish them. Therefore, the state-of-the-art image categorization methods do not perform sufficiently on insect images. In this paper, we propose an insect species recognition method based on class specific sparse representation. On obtaining the vector representation of image via sparse coding of patches, an SVM classifier is used to classify the image into species. We propose two class specific sparse representation methods under weakly supervised learning to discriminate insect species which have substantial similarity to each other. Experimental results show that the proposed methods perform well in insect species recognition and outperform the state-of-the-art methods on generic image categorization.

1 Introduction

Insect species recognition is widely applied in agriculture, ecology, and environmental science. Comparing to face recognition and generic object categorization, insect species recognition needs more expert knowledge. That means without some professional experience it is almost impossible for laymen to determine an insect category in the level of species. As a result, insect species recognition using computer vision methods is more and more required in application. The goal of our research is to develop a computer vision method which is convenient (without so much interaction such as alignment, rotation and segmentation), foolproof (not necessary to have expert knowledge), and inexpensive (only needing a PC, a digital camera and little human labor) to partly take place of expert entomologists in some area and help entomologists lighten heavy labor in their researches.

Besides its practical importance, automated recognition of insect species raises many fundamental computer vision challenges. Most insects are composed of several sub-parts

© 2010. The copyright of this document resides with its authors.
It may be distributed unchanged freely in print or electronic forms.
BMVC 2010 doi:10.5244/C.24.108

(legs, antennae, tails, wing pads, etc.) and many degrees of freedom. Some species are distinctive; others are very difficult to identify [4]. To develop a method which is invariant to pose, size, and orientation is the challenge we have to face in insect species recognition and many other computer vision applications.

Our bases construction methods are motivated by the sparse coding spatial pyramid matching (ScSPM) model for generic image categorization by Yang *et al*. [13], sparse representation based classification (SRC) algorithm for face recognition by Wright *et al*. [11] and image restoration method by Mairal *et al*. [8]. All of the three methods are based on sparse representation which is very popular in computer vision, image processing and machine learning. The term "sparse representation" refers to an expression of the input signal as a linear combination of base elements in which many of the coefficients are zero [11]. Comparing to [13], [11] and [8], we calculate bases of each class to obtain the class specific sparse representations in the strategies of minimal reconstruction residual and sparsity of local features. Experiments on insect species dataset and generic image categorization dataset show the validity of our approach.

The remainder of this paper is organized as follows. In Section 2 we will talk about some related works. Section 3 details class specific sparse representation methods which is an extension of the traditional sparse coding. Section 4 presents the application problem of insect species recognition. In Section 5, we will give the experiment results on our Tephritidae dataset and Caltech 101 dataset [7]. Finally, Section 6 concludes our paper.

2 Related works

We divide our discussion of related works into two parts. First, we review related works in insect recognition systems. Then we discuss some related works in generic object categorization and other areas.

2.1 Insect recognition systems

Over the years many works have been done on automated insect recognition. Species identification, automated and web accessible (SPIDA-web) [3] is an automated spider[1] species recognition system that applies neural networks to wavelet encoded images. Digital automated identification system (DAISY) [9] based on eigen-images is applied to several families of insects. The automated bee identification system (ABIS) [1] takes geometric features from a photograph of a bee's forewing and uses SVM for classifying. All of the three approaches require manual manipulations and interactions. Larios *et al*. [4] developed an insect identification system which combines PCBR detector, SIFT descriptor, bag-of-features model and logistic model tree as learning algorithm. This system doesn't require any interaction (weakly supervised) and gets good result (80%+) for application. However, compared to our dataset (20 species and 3 to 20 images per species) their dataset has only 4 species and 124 to 463 images per species which is much easier for training and classifying.

[1] In the view of biology, spiders are not belong to Insecta. However spiders and insects have much similarity and are both belong to Arthropoda. So we introduce the spider recognition system together with the other insect recognition systems.

2.2 Generic object categorization and others

In the past decade several efficient approaches to generic object categorization have appeared. Among them, bag-of-features (BOF) [2], spatial pyramid matching (SPM) [5] and sparse coding spatial pyramid matching (ScSPM) [13] are the most popular methods based on local features. These methods work by partitioning the image into small patches, computing histograms or sparse codes and taking single or multiple layers' pooling to represent an image by a vector. Then, any training method such as SVM could be used for classifying. This framework is simple and computationally efficient so our work is also based on this framework. Another tendency is that sparse representation is more and more widely used in object categorization and other areas such as face recognition. Sparse representation is powerful due to the fact that most important kinds of signal (audio or images) have naturally sparse representations with respect to fixed bases [12]. Wright *et al.* [11] proposed a robust and highly accurate face recognition method based on sparse representation. However their face dataset are all cropped and normalized so that the overcomplete dictionary are the training images themselves. But this is not the case in our problem. We wish to develop a weakly supervised approach without any manual interaction so we benefit from the patch-wise residual idea from image restoration method by Mairal *et al.* [8].

3 Class specific sparse representation methods

We first introduce some conceptions of sparse representation. Then we detail our class specific bases construction and sparse coding methods.

3.1 Conception

The goal of sparse representation (coding) is to represent an input vector approximately as a linear combination of a small number of basis vectors (column of the codebook or dictionary). These basis vectors can capture high-level patterns in the input data [6]. Let $X = [x_1, x_2, \ldots, x_N] \in \mathbb{R}^{D\times N}$ be the input matrix (each column is an input vector), let $B = [b_1, b_2, \ldots, b_K] \in \mathbb{R}^{D\times K}$ be the basis matrix (each column is a basis vector), and let $S = [s_1, s_2, \ldots, s_N] \in \mathbb{R}^{K\times N}$ be the coefficient matrix (each column is a coefficient vector). D is the dimension of input vectors, N is number of input vectors, and K is the number of bases. Then, the optimization problem above can be formulated as:

$$\begin{aligned} &\min_{B,S} \sum_{n=1}^{N} \|x_n - Bs_n\|_2^2 + \lambda \|s_n\|_1 \\ &s.t. \;\; \|b_k\|_2 \le c, \;\; \forall k = 1,2,\ldots,K \end{aligned} \tag{1}$$

Where $\|\cdot\|_2$ means Euclidean ($L2$) norm and $\|\cdot\|_1$ means $L1$ norm. The constraint for bases: $\|b_k\|_2 \le c, \;\; \forall k = 1,2,\ldots,K$ is necessary because we can respectively multiply and divide an infinitive large constant to B and s_n which keeps $\sum_{n=1}^{N} \|x_n - Bs_n\|^2$ unchanged while making s_n approach 0. However this is a trivial solution.

3.2 Class specific bases construction and sparse coding methods

Our bases construction method is based on the work of Yang *et al.* [13]. However their method is more suitable for generic object image datasets which have more distinction be-

tween classes than that of insect species. So motivated by the work of Wright *et al*. [11], we adopt a class specific bases construction strategy. The holistic optimization problem can be formulated as:

$$\begin{aligned} &\min_{B_i,S^{(i)}} \sum_{n_i=1}^{N_i} \|x_{n_i}^{(i)} - B_i s_{n_i}^{(i)}\|_2^2 + \lambda \|s_{n_i}^{(i)}\|_1, \quad for \;\; i = 1,2,\ldots,C \\ &s.t. \;\; \|b_{i,k_i}\|_2 \leq c, \;\; \forall k_i = 1,2,\ldots,K_i \end{aligned} \tag{2}$$

For each class we calculate its own basis matrix by an iterative process. Firstly we randomly initialize basis matrix B_i to calculate the new sparse codes $s_{n_i}^{(i)}$ for input vector $x_{n_i}^{(i)}$ for each class i :

$$\min_{s_{n_i}^{(i)}} \|x_{n_i}^{(i)} - B_i s_{n_i}^{(i)}\|_2^2 + \lambda \|s_{n_i}^{(i)}\|_1 \tag{3}$$

Then we fix sparse codes S and solve the following optimization problem with constraint:

$$\begin{aligned} &\min_{B_i} \sum_{n_i=1}^{N_i} \|x_{n_i}^{(i)} - B_i s_{n_i}^{(i)}\|_2^2 \\ &s.t. \;\; \|b_{i,k_i}\|_2 \leq c, \;\; \forall k_i = 1,2,\ldots,K_i \end{aligned} \tag{4}$$

Lee *et al*. [6] has developed an efficient algorithm for solving this problem. We define the basis matrix of each class as: $B_i \doteq [b_{i,1}, b_{i,2}, \ldots, b_{i,K_i}] \in \mathbb{R}^{D \times K_i}$ and C is the number of classes. Wright *et al*. [11] combine the C basis matrices together to make a new matrix: $B \doteq [B_1, B_2, \ldots, B_C] = [b_{i,1}, b_{i,2}, \ldots, b_{C,K_C}]$, however, we keep the C basis matrices separately because of our local feature extraction strategy. Our method is under the assumption that the basis matrix calculated from a class takes more discriminative information and is more precise to reconstruct a new feature vector from the same class. Then we can take advantage of the reconstruction residual introduced in [8].

Then, for any new input vector x_{new}, we can get C coefficient vectors (sparse codes) $s_{new}^{(i)}$ respectively to each basis matrix by solving the optimization problem:

$$\min_{s_{new}^{(i)}} \|x_{new} - B_i s_{new}^{(i)}\|_2^2 + \lambda \|s_{new}^{(i)}\|_1, \quad for \;\; i = 1,2,\ldots,C \tag{5}$$

Feature-sign search algorithm [6]is so efficient to solve this problem. Then we proposed two strategies to concatenate the C coefficient vectors into one sparse vector to represent the original input vector. The first one we call it minimal residual class specific sparse representation (MRCSSR). That means we take the coefficient vector which minimizes the residual of reconstruction as its original value and other vector as zero.

$$\begin{aligned} p &= \arg\min_i \|x_{new} - B s_{new}^{(i)}\|_2 \\ s_{new} &= [\underbrace{0,0,\ldots,0}_{K_1}, \underbrace{0,0,\ldots,0}_{K_2}, \ldots, \underbrace{(s_{new}^{(p)})^T}_{Kp}, \ldots, \underbrace{0,0,\ldots,0}_{K_C}]^T \end{aligned} \tag{6}$$

The second strategy we call it sparsest class specific sparse representation (SCSSR). That means we take the coefficient vector which is sparsest to represent the original feature and other vector as zero. Here we take L_0 norm to evaluate the sparsity of the coefficient vectors.

$$\begin{aligned} p &= \arg\min_i \|s_{new}^{(i)}\|_0 \\ s_{new} &= [\underbrace{0,0,\ldots,0}_{K_1},\underbrace{0,0,\ldots,0}_{K_2},\ldots,\underbrace{(s_{new}^{(p)})^T}_{Kp},\ldots,\underbrace{0,0,\ldots,0}_{K_C}]^T \end{aligned} \tag{7}$$

After calculating the sparse representation of each input vectors, any pooling method such as averaging or max pooling [13] can combine these sparse codes of input vectors belonging to the same sample together to obtain the final feature vectors. Then any learning method such as neural networks or SVM is competent for the recognition task.

4 Application of insect species recognition

Tephritidae (fruit fly) is a family of insect which contains about 500 genera and about 4200 species [10]. A given species is harmful to specific one or two plants. However different species appear too similar (as shown in the fist row of Fig.2) to be recognized by laymen without any entomology knowledge. So it is urgent and economical to solve this problem by computer vision methods. The goal of our approach is to recognize different Tephritidae species without any interaction such as cropping, rotating or normalizing in both training and testing stages.

Our scheme framework is shown in Fig.1. A dataset is divided into two parts: training dataset and testing dataset. For both datasets, a sample is an image of Tephritidae (making use of species labels in training dataset and not making use of the labels in testing dataset). We transform all the images into gray scale and extract SIFT features of patches by densely sampled from each image. So an image sample can be represented by a set of SIFT features. After that we obtain C basis matrices $B_1, B_2, \ldots, B_C$ from the training dataset using Eq.(3) and Eq.(4). And by Eq.(5) each SIFT feature can be translated into C sparse vectors $s_1, s_2, \ldots, s_C$ .Then we concatenate the C sparse vectors into one vector by Eq.(6) or Eq.(7). Subsequently we calculate the pooling function of all the vectors of patches in the same image. This spatial pyramid pooled vector is the final representation of an image sample which can be used by any machine learning method for classification. For a testing image represented by a set of SIFT features, it can also be translated into a sparse vector by the basis matrices calculated aforementioned and we can use the classifier we have learned to determine the image label.

Our Tephritidae dataset is composed of 3 genera and 20 species. Each specimen is taken one photograph respectively of its whole body, head, thorax, abdomen and wing (as shown in Fig.2). So we divide the whole dataset into 5 sub-dataset according to different part of specimen. Because the wings are crisp and vulnerable in the preservation, the photographs of wings are less than that of the other parts. Considering the case that there is only one photograph of a specimen in some species, we have no idea to divide the data of these species into training and testing subsets. So we discard these photographs to obtain a dataset appropriate for experiment. Table. 1 shows the number of species and photographs of the 5 sub-datasets.

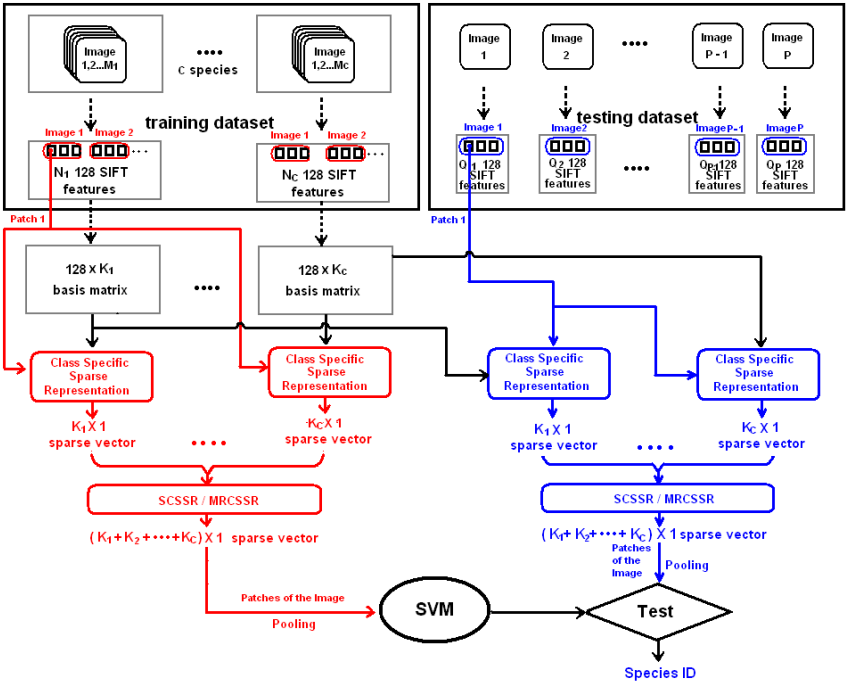

Figure 1: Our insect species recognition scheme framework. N_i is the number of training local features in species i, Q_j is the number of local features in testing image j and K_m is the number of bases in species m.

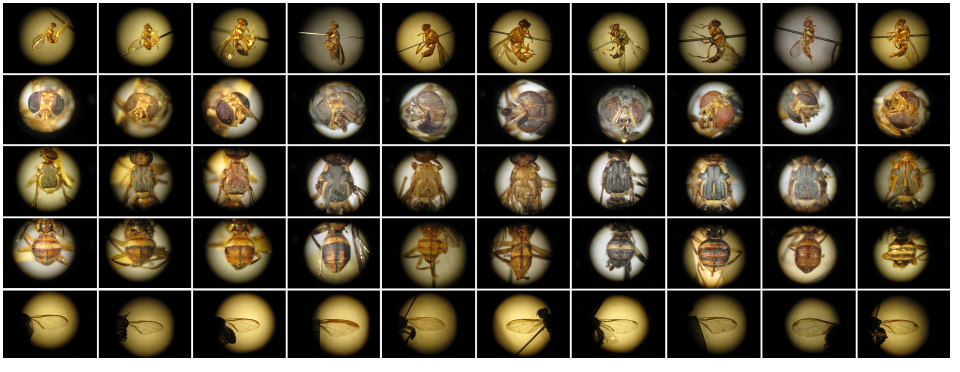

Figure 2: Example of our Tephritidae dataset: each column is corresponding to one species and the rows are respectively whole body, head, thorax, abdomen and wing photographs of the corresponding species taken by a microscope camera.

Sub-dataset	Whole	Head	Thorax	Abdomen	Wing
Species	19	19	20	17	14
Images	152	143	151	144	103

Table 1: Number of species and images in each sub-dataset.

Sub-dataset	1 training sample			2 training samples		
	1 layer	2 layers	3 layers	1 layer	2 layers	3 layers
Whole	**63.11 ± 0.37**	46.47 ± 0.20	39.07 ± 0.53	**77.91 ± 0.16**	65.94 ± 0.47	60.49 ± 0.28
Head	**67.99 ± 0.19**	64.41 ± 0.24	55.91 ± 0.45	**80.15 ± 0.18**	78.77 ± 0.27	75.61 ± 0.26
Thorax	**77.22 ± 0.18**	74.88 ± 0.35	70.06 ± 0.43	**83.35 ± 0.10**	82.20 ± 0.13	79.13 ± 0.16
Abdomen	**70.73 ± 0.27**	66.61 ± 0.36	59.47 ± 0.38	**80.80 ± 0.14**	78.22 ± 0.35	74.50 ± 0.13
Wing	**56.46 ± 0.68**	48.94 ± 0.25	49.52 ± 0.38	**71.56 ± 0.49**	65.71 ± 0.36	60.76 ± 0.29

Table 2: Recognition rate (%) with SCSSR method.

5 Experiments and results

5.1 Experiment configuration

In the experiments, we evaluate our two class specific sparse representation methods on the Tephritidae dataset and the generic Caltech101 dataset. As in [13] we extract SIFT descriptor from 16×16 pixel patches which are densely sampled from each image on a grid with a stepsize of 8 pixels. We use multi-layer Max-pooling ScSPM [13] and linear SVM for training the classifier and evaluate our result by 10-fold cross validation.

5.2 Results

Firstly, we show our results on the Tephritidae dataset. Because some species of the dataset has only 3 photographs, we can obtain results by taking 1 or 2 samples from each species for training and the others for testing. We set the number of layers of spatial pyramid as 1, 2 or 3. The number of bases constructed in each species is fixed to 256. The result of our sparsest class specific sparse representation (SCSSR) method and minimal residual class specific sparse representation (MRCSSR) method are respectively shown in Table 2 and Table 3.

Through the results we find that when the number of layers of spatial pyramid increases, however, the recognition rate almost synchronously decreases. This may be caused by the

Sub-dataset	1 training sample			2 training samples		
	1 layer	2 layers	3 layers	1 layer	2 layers	3 layers
Whole	**67.41 ± 0.48**	60.60 ± 0.35	49.30 ± 0.45	**78.84 ± 0.24**	73.42 ± 0.57	54.38 ± 0.36
Head	**71.07 ± 0.17**	64.39 ± 0.19	61.67 ± 0.47	**83.53 ± 0.20**	81.66 ± 0.26	70.20 ± 0.26
Thorax	**80.14 ± 0.26**	77.31 ± 0.18	74.37 ± 0.18	88.50 ± 0.12	**88.74 ± 0.17**	80.18 ± 0.24
Abdomen	**76.29 ± 0.26**	73.11 ± 0.38	66.27 ± 0.45	**85.68 ± 0.23**	83.57 ± 0.26	70.86 ± 0.36
Wing	**62.20 ± 0.51**	54.00 ± 0.41	54.60 ± 0.59	73.93 ± 0.51	**74.29 ± 0.41**	60.75 ± 0.35

Table 3: Recognition rate (%) with MRCSSR method.

Sub-dataset	Whole	Head	Thorax	Abdomen	Wing
SPM	24.73 ± 0.27	21.68 ± 0.47	27.66 ± 0.38	26.03 ± 0.62	26.15 ± 0.53
ScSPM	25.81 ± 0.48	23.52 ± 0.51	31.32 ± 0.40	33.04 ± 0.65	38.55 ± 0.63
SCSSR	77.91 ± 0.16	80.15 ± 0.18	83.35 ± 0.10	80.80 ± 0.14	71.56 ± 0.49
MRCSSR	**78.84 ± 0.24**	**83.53 ± 0.20**	**88.50 ± 0.12**	**85.68 ± 0.23**	**73.93 ± 0.51**

Table 4: Recognition rate (%) comparison of different methods on each sub-dataset.

Number of training samples	15	30
SPM	54.28 ±0.56	64.53 ±0.72
ScSPM	65.46 ±0.69	73.71 ±0.53
SCSSR	68.12 ±0.73	74.21 ±0.82
MRCSSR	**69.38 ±0.54**	**75.15 ±0.66**

Table 5: Recognition rate (%) comparison of different methods on Caltech101 dataset.

arbitrary orientation of the insect located in the images. As far as we know, the SPM method can not solve the rotation invariance problem. So if the insects in the dataset have severely varied orientations, multi-layer SPM does not work better than single-layer SPM (which degenerates to bag-of-feature). According to the result, we can also deduce this argument. As shown in Fig.1, the insect images in Whole Body and Head sub-dataset is more variant in orientations. As a matter of fact, by increasing the number of layers the results of these two sub-dataset deteriorate severely while the results of stable orientation sub-datasets such as Thorax's deteriorate more mildly. We have to say another reason that the Thorax sub-dataset gets the best result is that the Thorax part is more salient and takes more discriminative information to distinguish different species. We can get this conclusion by comparing different sub-datsets in Fig.2. Further more, we find that the MRCSSR method is superior to the SCSSR method. This gives us some cue that reconstruction residual may take more discriminative information in recognition.

Secondly, we compare our methods with ScSPM [13]and traditional SPM [8] on our dataset. To be fair, we set the number of the holistic bases in ScSPM and the number of cluster center in SPM to $256 \times C$, where C is number of species in each sub-dataset. We set the number of layer to 1 and the number of training samples to 2. The result is shown in Table 6. According to the result our two methods remarkably outperforms the other methods. It may be because that the class specific methods can discover the most discriminative representation and are more suitable for our Tephritidae dataset.

Finally, we compare our methods with ScSPM and traditional SPM on Caltech101 dataset. The result in Table 5 shows that our methods outperform the others. The reason that the recognition rates are not promoted remarkably by our methods is that we limit the number of bases to 20 of each category in order to get a moderate length of the concatenated sparse vector. Too little bases may cause the loss of expressivity to represent the SIFT features of batches.

6 Conclusions and future works

In this paper we proposed two class specific sparse representation methods for insect species recognition. These methods use class specific sparse vectors instead of traditional sparse coding vectors for insect image patches. Our experiments on the Tephritidae dataset and Caltech101 dataset demonstrated the effectiveness of our methods. We believe that constructing a basis matrix for each class will take more discriminative information into the final sparse representation and both the minimal residual and the sparsest strategy remove some noise among other similar classes and remain the information which is utmost expressive for the true classes.

Our further works will focus on the following three directions: 1. The SPM pooling methods can not solve the case of orientation variety. So we are interested in some rotation invariance methods to make the sparse representation more robust. 2. Now our experiments are conducted separately on the five different sub-datasets. We suppose that a user may take photographs of several sub-parts of an insect. So how to combine all recognition outputs together to get a better result is mainly our task in next stage. 3. In spite of our class specific sparse representation methods earn good results and take some discriminative information from the separation of basis construction of different classes, there is no explicit discriminative terms in the optimization function to calculate basis matrices. If the optimizing process contains some explicit discriminative terms we suppose that the recognition result will be much better.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (NSFC) under grant no.60825301 and the National Laboratory of Pattern Recognition Program at the Institute of Automation of the Chinese Academy of Sciences and National Science Fund for Fostering Talents in Basic Research of China (Special subjects in animal taxonomy, NSFC-J0630964/J0109).

References

[1] T. Arbuckle, S. Schroder, V. Steinhage, and D. Wittmann. Biodiversity informatics in action: identification and monitoring of bee species using abis. In *Proceedings of the 15th International Symposium Informatics for Environmental Protection*, 2001.

[2] G Csurka, C. Dance, L. Fan, and C. Williamowski, J.and Bray. Visual categorization with bags of keypoints. In *Workshop on Statistical Learning in Computer Vision, ECCV*, pages 1–22, 2004.

[3] M. Do, J. Harp, and K. Norris. A test of a pattern recognition system for identification of spiders. *Bull. Entomol. Res*, 89(3):217–224, 1999.

[4] N. Larios, H.L. Deng, W. Zhang, M. Sarpola, J. Yuen, R. Paasch, A. Moldenke, D.A. Lytle, S.R. Correa, E.N. Mortensen, L.G. Shapiro, and T.G. Dietterich. Automated insect identification through concatenated histograms of local appearance features: feature vector generation and region detection for deformable objects. *Machine Vision and Applications*, 19:105–123, 2008.

[5] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *CVPR*, pages 2169–2178, 2006.

[6] H. Lee, A. Battle, R. Raina, and A. Y. Ng. Efficient sparse coding algorithms. In *NIPS*, pages 801–808, 2006.

[7] F.F. Li, R. Fergus, and P. Perona. Learning generative visual models from few training examples:an incremental bayesian approach tested on 101 object categories. In *CVPR Workshop on Generative-Model Based Vision*, 2004.

[8] J. Mairal, M. Elad, and G. Sapiro. Sparse representation for color image restoration. *IEEE Trans. IP*, 17(1):53–69, 2008.

[9] M.A. O'Neill, I.D. Gauld, K.J. Gaston, and P.J.D. Weeks. Daisy: An automated invertebrate identification system using holistic vision techniques. In *Proceeding of the Inaugural Meeting BioNET-International Group for Computer-Aided Taxonomy*, 2000.

[10] X.J. Wang. The fruit flies (diptera: Tephritidae) of the east asian region. *Acta Zootaxon Sinica*, 21(Supplement):1–338, 1996.

[11] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Trans. PAMI*, 31(2):210–227, 2009.

[12] J. Wright, Y. Ma, J. Mairaly, G. Sapiro, T. Huang, and S. Yan. Sparse representation for computer vision and pattern recognition. In *PROCEEDINGS OF IEEE*, MARCH 2009.

[13] J. Yang, K. Yu, Y.H. Gong, and T. Huang. Linear spatial pyramid matching using sparse coding for image classification. In *CVPR*, 2009.