Insect Species Recognition using Sparse Representation

An Lu Student
alu@nlpr.ia.ac.cn
Xinwen Hou Prof
xwhou@nlpr.ia.ac.cn
Xiaolin Chen Prof
xchen@ioz.ac.cn
Chenglin Liu Prof
liucl@nlpr.ia.ac.cn

Insect species recognition is a typical application of image categorization and object recognition. 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.

Our work is motivated by the sparse coding spatial pyramid matching (ScSPM) model for generic image categorization [3], sparse representation based classification (SRC) algorithm for face recognition [2] and image restoration method [1]. All of the three methods are based on sparse representation. Comparing to these methods, 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.

Our bases construction method is based on [3]. However, their method is more suitable for generic object image datasets which have more distinction between classes than that of insect species. So motivated by the work of [2], we adopt a class specific bases construction strategy. The holistic optimization problem can be formulated as:

$$\min_{B_i, S_i} \sum_{n=1}^{N} \left\| x_n^{(i)} - B_i s_n^{(i)} \right\|_2^2 + \lambda \left\| s_n^{(i)} \right\|_1 \quad \text{for } i = 1, 2, \ldots, C$$

(1)

s.t. \left\| B_i k \right\|_2 \leq c, \quad \forall k = 1, 2, \ldots, K_i

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 $s_n^{(i)}$ for input vector $x_n^{(i)}$ for each class $i$:

$$\min_{s_n^{(i)}} \left\| x_n^{(i)} - B_i s_n^{(i)} \right\|_2^2 + \lambda \left\| s_n^{(i)} \right\|_1$$

(2)

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

$$\min_{B_i} \sum_{n=1}^{N} \left\| x_n^{(i)} - B_i s_n^{(i)} \right\|_2^2$$

s.t. \left\| B_i k \right\|_2 \leq c, \quad \forall k = 1, 2, \ldots, K_i

At the same time, 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:

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

(4)

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.

$$p = \arg \min_{i} \left\| x_{new} - B_i s_{new}^{(i)} \right\|_2$$

$$s_{new} = \left[ 0, 0, \ldots, 0, 0, \ldots, 0 \right]^T + \left[ \left( s_{new}^{(p)} \right)^T \right]_{K_p}$$

$$K_1 K_2 \ldots K_p$$

(5)

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.

$$p = \arg \min_{i} \left\| s_{new}^{(i)} \right\|_0$$

$$s_{new} = \left[ 0, 0, \ldots, 0, 0, \ldots, 0 \right]^T + \left[ \left( s_{new}^{(p)} \right)^T \right]_{K_p}$$

$$K_1 \ldots K_p$$

(6)

After calculating the sparse representation of each input vectors, any pooling method such as averaging or max pooling [3] 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.

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.1). So we divide the whole dataset into 5 sub-dataset according to different part of specimen. Table 1 shows the number of species and photographs of the 5 sub-datasets.

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.