

# Multi-kernel dictionary learning for classifying maize varieties

Hua Zhu<sup>1,2</sup>, Jun Yue<sup>1\*</sup>, Zhenbo Li<sup>2</sup>, Zhiwang Zhang<sup>1</sup>

(1. College of Information and Electrical Engineering, Ludong University, Yantai 264025, Shandong, China;

2. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China)

**Abstract:** The automatic classification and identification of maize varieties is one of the important research contents in agriculture. A multi-kernel maize varieties classification approach was proposed in this paper in order to improve the recognition rate of maize varieties. In this approach, four kinds of maize varieties were selected, in each variety 200 grains were selected randomly as the samples, and in each sample 160 grains were taken as the training samples randomly; the characteristics of maize grain were extracted as the typical characteristics to distinguish maize varieties, by which the dictionary required by K-SVD was constructed; for the test samples, the feature-matrixes were extracted by dimension reduction method which were mapped to the high-dimension space by multi-kernel function mapping. The high-dimension characteristic matrixes were trained by K-SVD method and the corresponding feature dictionary was obtained respectively. Finally, the test samples representing were trained and classified by  $l_{2,1}$  minimization sparse coefficient. The experiment results showed that recognition rate was improved obviously through this approach, and the poor-effect to maize variety identification from partial occlusion can be eliminated effectively.

**Keywords:** multi-kernel, sparse representation, dictionary learning, maize classification

**DOI:** 10.25165/j.ijabe.20181103.3091

**Citation:** Zhu H, Yue J, Li Z B, Zhang Z W. Multi-kernel dictionary learning for classifying maize varieties. Int J Agric & Biol Eng, 2018; 11(3): 183–189.

## 1 Introduction

Maize is one of the main grain crops in the world, and it is also the main grain and economic crop in China. As the main grain crop, maize plays an important role in meeting people's dietary requirements. By maize varieties identification technology, maize classifying automatically and high-quality maize selecting effectively can be realized to promote the process of maize production, processing and export trade. At present, the classification of maize mainly depends on manual evaluation of its shape, color and other aspects. It has the disadvantages of strong subjectivity and low efficiency, which increases the uncertainty of maize varieties classification. With the continuous development of computer technology, machine vision technology is used more and more widely in quality inspection and classification of agricultural products, and it has effectively improved the efficiency of agricultural production. Machine vision instead of manual identification has the following advantages: (1) Multi-parameters measurement, comprehensive evaluation and classification; (2) Reduce human subjective factors and realize classification automatically; (3) Reduce inspecting error and improve accuracy.

There are many different kinds of maize with big different morphology, color etc., which lead to different maize classification methods in different countries, and classification methods are

mostly focused on machine vision research. In Europe and America, the identification and classification by machine vision technology develop rapidly. Thinking about grain surface defects<sup>[1-3]</sup>, color difference, size difference, roughness and texture difference, computing by machine vision intelligent algorithms, by which the classification effect is achieved and human subjective factors are eliminated successfully. In the domestic research, progress has been made in the detection and identification of maize varieties using machine vision technology. Grains collecting and processing are carried out by machine vision, characteristic parameters of maize are chosen and optimized<sup>[4-9]</sup> by a series of algorithm such as artificial neural network model, support vector machines etc. Some scholars go on identification and classification by the combining infrared spectroscopy and intelligent algorithm. However, the research time to machine vision for agricultural products classification is not long, and there are still many deficiencies. The correlation between characteristic parameters of agricultural products and the occluded factor is not established, the optimization of feature parameters is still inconclusive, which need further study.

In this paper, a new approach to identifying and classifying maize varieties is proposed. The perfect maize grain and the occluded grain can be recognized and classified by constructing the K-SVD model and improving the constructed K-SVD model respectively. The validity is proved by experiments.

## 2 Related works

### 2.1 Sparse representation

In recent years, sparse representation theory<sup>[7]</sup> has been used more and more in image processing, including image restoration, image denoising, image recognition etc. mainly. Specially, Sparse representation is applied to face recognition. Sparse representation<sup>[10]</sup> means using fewer (sparse) non-zero elements to represent signals, making signal processing more simple.

**Received date:** 2017-10-30    **Accepted date:** 2017-12-17

**Biographies:** **Zhu Hua**, MS, research interests: signal and image processing, Email: m18253559378@163.com; **Li Zhenbo**, PhD, research interests: digital image processing and computer vision, Email: lizb@cau.edu.cn; **Zhang Zhiwang**, PhD, research interests: data mining, artificial intelligence and natural language, Email: zzwms@126.com.

**\*Corresponding author:** **Yue Jun**, PhD, Professor, research interests: large scale visual computing and big data. Mailing address: No.186 Hongqi Middle Road, Zhifu District, Yantai 264025, China. Tel: +86-13562559603, Email: yuejuncn@126.com.

The principle of sparse representation is to use the sparse non-zero elements in the training sample to represent the test samples to achieve the purpose of identification. Use a few atoms in over-complete dictionary to represent test sample, the original signal can be reconstructed by only a small of trained data in the training sample. This method can not only identify test sample effectively but also reject invalid test samples outside the training database effectively. For different classes of training samples, the test samples can be identified and classified by the combination of a small number of atoms in the training sample as long as the sample is enough. In performance, sparse representation which uses a small number of samples to build a training set achieves better results than artificial neural networks and support vector machines (SVM).

**2.2 Dictionary Learning**

SRC (Sparse Representation Classifier) change all the training images into a matrix and reconstructs them, so that the amount of calculation is larger in the case of larger training samples. Therefore, the training sample need to be pretreated—learning, which means obtaining the optimal base—dictionary.

From the sparse representation principle, the test sample  $y$  can be obtained by dictionary  $D$  multiplying sparse representation matrix  $\alpha$ . The quality of the dictionary depends on the error size under certain sparsity, that is, the smaller the error, the closer the dictionary closing the original training set. The purpose of training a dictionary is to reduce the representation error and approximate the dictionary with the strongest representation.

There are nonlinear data in dictionary atoms, and the kernel<sup>[11-12]</sup> provides a relatively simple way to handle nonlinear problems. In recent years, kernels have also been under the sparse representation domain whatever in sparse coding or dictionary learning<sup>[16,17]</sup>. Although the classification results have been improved by K-SVD<sup>[16]</sup> method, the algorithm requires a very large kernel matrix, which results in the storage and processing with high computational cost, and its use is also limited. In order to solve this problem, non-linear kernel dictionaries learning method<sup>[18]</sup> is proposed. Moreover, the dictionary learning process can be extended to nonlinear, using appropriate kernel functions to handle nonlinear structured data, while multi-kernel learning is an effective way to deal with this. The iterative algorithm for consistent learning of dictionary matrix and multi-kernel<sup>[20-22]</sup> functions is proposed under sparse representation framework in reference<sup>[19]</sup>. The reconstruction error of sparse coded data is minimized by optimizing the sum of basis functions weighted by a set of multi-kernel functions.

**3 Multi-kernel based maize identifying**

In this approach, a model for maize varieties recognition was established by image collecting, image preprocessing, feature parameter extracting, multi-kernel designing and dictionary establishing etc. Finally, the effectiveness of the method was proved by experiments.

The technical routes of this study is shown in Figure 1.

**3.1 Characteristic parameter of maize**

**3.1.1 Color characteristics of maize grain**

HIS color model is a frequently-used color model in digital image processing. The color characteristic of maize grain is extracted by HIS model. The color characteristic of maize grain is represented by the average value of chrominance component, saturation component and brightness component of grain color, which meets the requirement of maize varieties identification to

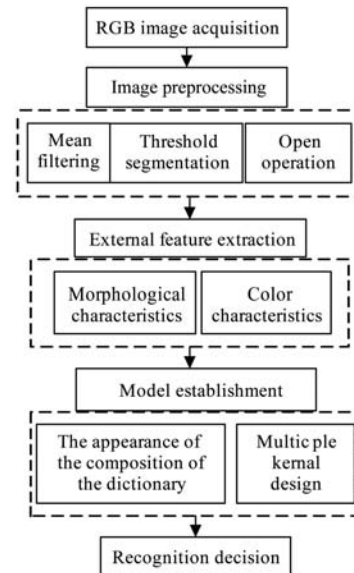


Figure 1 Technical routes

multi-parameter measurement of grain color characteristics. Functions are as follows:

$$f_{j,1}^i = \frac{1}{p} \sum_{k=1}^n I_k^H \tag{1}$$

$$f_{j,1}^i = \frac{1}{p} \sum_{k=1}^n I_k^S \tag{2}$$

$$f_{j,3}^i = \frac{1}{p} \sum_{k=1}^n I_k^I \tag{3}$$

where,  $f_{j,1}^i, f_{j,2}^i, f_{j,3}^i$  express the first, the second and the third characteristic values of  $j$  grain of  $i$  maize variety respectively,  $n$  is the number of pixels belong to the maize grain in the image,  $I_k^H, I_k^S, I_k^I$  express chrominance component, saturation component and brightness component of  $k$  point in grain image  $I$ .

**3.1.2 Shape characteristics of maize grain**

Shape difference is one of the important differences between different varieties of maize. The selection of morphological feature parameters is not uniform. It can be taken as feature parameters as long as the morphological differences can be distinguished effectively, and the parameters can be obtained quickly and conveniently. Area, rectangularity and elongation are taken as parameters. The morphological characteristics of maize grains are as follows.

Rectangularity:

$$f_{j,4}^i = \frac{A}{a * b} \tag{4}$$

Elongation:

$$f_{j,5}^i = \frac{b}{a} \tag{5}$$

Area:

$$f_{j,6}^i = A \tag{6}$$

where,  $A$  is the area of maize grain;  $a$  is the long axis, and  $b$  is the short axis.

**3.1.3 Appearance characteristic vector of maize**

To sum up, assume the kinds of samples are  $n$ , and  $m$  training samples for each kind, the appearance characteristic vector of the maize can be expressed as:

$$A^i = [f_1^i f_2^i \dots f_m^i] \tag{7}$$

where, the  $m$  column vectors form a space which reflects class  $I$  maize variety, and the dictionary matrix consisting of  $n$  classes training samples is as below.

$$A = [A^1 A^2 \dots A^n] \quad (8)$$

The row numbers of matrix  $A$  is the number of characteristic parameters of training samples, and the column numbers is the total number of training samples.

Forty grains are chosen as training samples for each maize variety, the characteristic vectors of these grains are arranged in order to form the following training sample matrix, that is, the dictionary:

$$\begin{aligned} A &= [f_1^1 \dots f_{40}^1 f_1^2 \dots f_{40}^2 \dots f_1^4 \dots f_{40}^4] \\ &= [A^1 A^2 \dots A^4] \end{aligned} \quad (9)$$

The matrix  $A$  express that dictionary consists of the characteristics of the four types of maize varieties, in which the numbers of row represent the number of extracted characters of maize and the numbers of column represent the total number of training samples.

### 3.2 The designing of multi-kernel based maize variety classification method

#### 3.2.1 The designing of multi-kernel function

The reduced dimension method PCA is used to training samples to obtain linearly non-separable data. Therefore, in order to make the trained dictionary represent the entire training set better, the linearly non-separable data are taken into account: Data after PCA dimension reduction are mapped into high-dimensional space by kernel function, that is feature space. In the feature space, these data will become into linearly separable, thus which can be processed simplistically.

With the development of SVM theory, people began to pay more attention to the kernel method<sup>[23]</sup>, which is an effective method for solving non-linear problems. However, the kernel function with a single kernel cannot meet the demand for complex applications, so the method of some kernels combination is proposed, and the multi-kernel learning method has become a new hotspot in kernel machine learning.

Multiple kernel learning method selects the corresponding kernel function of each characteristic among multi-features data. Due to the multi-kernel learning is based on multi-characteristics, so its data descriptive capability is better than the method with the optimal description of a single data feature. Thus, multi-kernel learning can achieve better learning performance by optimizing the weight coefficient<sup>[25]</sup> of kernel function. Multi-kernel learning not only retains the excellent nonlinear mapping characteristics of kernel function which maximize the advantage of different kernel functions, but also show the possibility that induce the classical methods with nonlinear feature extraction capability after kernel extension into a unified framework.

In this paper, Gauss kernel function and polynomial kernel function are used to form a multi-kernel function

$$k(x_i, x_j) = \eta_1 k_1(x_i, x_j) + \eta_2 k_2(x_i, x_j) \quad (10)$$

where,  $\eta_1 + \eta_2 = 1$ .

#### 3.2.2 Multi-kernel K-SVD dictionary training

The goal of this part of work is to train a corn kernel dictionary  $D = [d_1, d_2, \dots, d_k]$  in feature space  $F$ , which can be realized by solving the following optimization problem:

$$\begin{aligned} &\arg \min_{D, X} \|\Phi(Y) - DX\|_F^2 \\ &\text{s.t. } \|x_i\|_0 \leq T_0, \|d_j\|_2 = 1, \forall i, j \end{aligned} \quad (11)$$

where,  $\Phi(Y) = [\Phi(y_1), \Phi(y_2), \dots, \Phi(y_N)]$  is a matrix,  $\Phi(y_i)$  is  $D \in {}^{n \times K}, X \in {}^{K \times N}$

corresponding to the sparse vector  $x_i$ . is written as follows:

$$\begin{aligned} &\arg \min_{A, X} \|\Phi(Y) - DAX\|_F^2 \\ &\text{s.t. } \|x_i\|_0 \leq T_0, \forall i = 1 \end{aligned} \quad (12)$$

Kernel learning consists of two phases: sparse coding phase and dictionary updating phase.

1) Sparse coding: Assume  $\Phi(Y)$  keeps invariant and  $X$  is computed. The Equation (12) can be written as:

$$\begin{aligned} &\|\Phi(Y) - \Phi(Y)AX\|_F^2 \\ &= \sum_{j=1}^N \|\Phi(y_j) - \Phi(Y)Ax_j\| \end{aligned} \quad (13)$$

KOMP is adopted to solve Equation (16)

The signal  $z \in \mathbb{R}^n$  is given and  $A$  is used to represent the core dictionary, the sparse combination of dictionary atoms approximate  $z$  in feature space is looked for:

$$\begin{aligned} \Phi(Z) &= \Phi(Y)AX_s + r_s \\ &= \Phi(Y)AX_s + r_s \end{aligned} \quad (14)$$

where,  $v_s = AX_s \in \mathbb{R}^n$  is an intermediate variable and  $r_s$  is the current residuals.

$I_s$  is used to represent the selected atomic index, and the residuals are projected to the remaining dictionary atomic  $d_i = \Phi(Y)a_i$ , where  $i \notin I_s$  is the column  $i$  of  $A$  and  $T_i$  represents the amount of residuals projected on  $d_i$ , which is calculated by (15)

$$\begin{aligned} T_i &= r_s (\Phi(Y)a_i)^T \\ &= (\Phi(Z) - \Phi(Y)v_s)^T (\Phi(Y)a_i) \\ &= (K(z, Y) - v_s^T K(Y, Y))a_i \end{aligned} \quad (15)$$

where

$$K(Z, Y) = [k(z, y_1), k(z, y_2), \dots, k(z, y_N)]$$

A new dictionary atom is selected in the remaining set of the largest projection coefficients, which guarantees the maximum reduction of the approximation error.

To update the coefficients of the  $x$  corresponding to the selected index in  $I_s$ . Assume  $X_s$  represents the vector which removes all the coefficients belong to  $I_s$ . Similarly, let  $A_{I_s}$  indicate a set of letter atoms whose indexes come from the set  $I_s$ . Then,  $X_s$  are obtained by projecting the signal  $\Phi(Z)$  onto the subspace spanned by the selected dictionary atoms.  $\Phi(Y)A_{I_s}$

Where

$$(\Phi(Y)A_{I_s})^T (\Phi(Y)A_{I_s}) = A_{I_s}^T$$

is positive, since  $K(Y, Y)$  is semi-positive, the projection coefficients are as follows:

$$\begin{aligned} X_s &= ((\Phi(Y)A_{I_s})^T (\Phi(Y)A_{I_s}))^{-1} (\Phi(Y)A_{I_s})^T \Phi(Z) \\ &= (A_{I_s}^T k(Y, Y)A_{I_s})^{-1} (k(Z, Y)A_{I_s})^T \end{aligned} \quad (16)$$

2) Dictionary updating: The K-SVD method is adopted to update dictionaries.

Let  $\alpha_k$  and  $x_T^j$  denote the column  $k$  from  $A$  and the row  $j$  from  $x$  respectively.

$$\|\Phi(Y) - \Phi(Y)AX\|_F^2 \quad (17)$$

Equation (17) can be rewritten as

$$\begin{aligned} & \left\| \Phi(Y) - \Phi(Y) \sum_{j=1}^k \alpha_j x_T^j \right\|_F^2 \\ &= \left\| \Phi(Y) \left( 1 - \sum_{j \neq k} \alpha_j x_T^j \right) - \Phi(Y) (\alpha_k x_T^k) \right\|_F^2 \\ &= \left\| \Phi(Y) E_k - \Phi(Y) M_k \right\|_F^2 \end{aligned} \quad (18)$$

where

$$E_k = \left( 1 - \sum_{j \neq k} \alpha_j x_T^j \right), M_k = \alpha_k x_T^k$$

where,  $\Phi(Y)E_k$  is the error between the approximate signal and the real signal when the dictionary atom  $k$  is removed and  $\Phi(Y)M_k$  is the contribution of the dictionary atom  $k$  to the estimated signal.

Define  $\omega_k$  as

$$\omega_k = \{i | 1 \leq i \leq N, X_T^k \neq 0\}$$

Note that,  $\Omega_k$  is a matrix of size  $N \times |\omega_k|$ , where  $(\omega_k(i), i)$  is 1, the other is zero. When it multiplied by  $\Omega_k$ , all zeros within row vector  $X_T^k$  will be discarded. The column reduction matrix is obtained as follows:

$$\begin{aligned} E_k^R &= E_k \Omega_k; \\ M_k^R &= M_k \Omega_k \end{aligned} \quad (19)$$

SVD was used to decompose Equation (20)

$$(E_k^R)^T K(Y, Y) (E_k^R) = V \Delta V^T \quad (20)$$

### 3.2.3 Maize classification algorithm

The maize classification algorithm flowchart is as follows:

1) Input samples: 20% data are selected randomly from each corn variety sample as test sample, the rest of the samples is taken as training sample.

2) Dimension reduction: Two samples are characterized by extracting feature and reduced dimension by PCA, and the matrix  $A_1, A_2, A_3, A_4$  and vector  $y$  are obtained respectively.

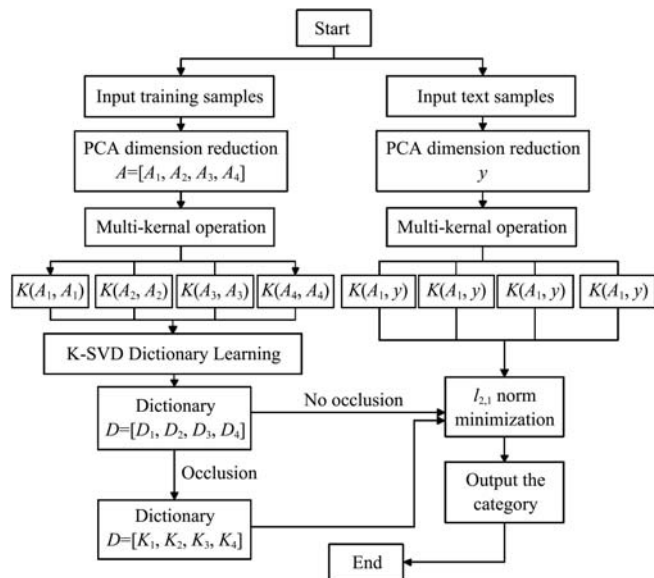


Figure 2 Maize classification algorithm flowchart

3) Multi-kernel operation:  $A_1, A_2, A_3, A_4, y$  is obtained which are linearly separable, that is, those non-linear separable data that can also reflect the image features are ignored, which is a loss to the information of the image and reduce the recognition rate. Therefore, the obtained matrix  $A_1, A_2, A_3, A_4$  and vector  $y$  are added

a kernel function to raise dimensions:

$$\begin{aligned} \Phi(\bullet) &: A_1 \rightarrow \Phi(A_1), \\ A_2 &\rightarrow \Phi(A_2), \\ A_3 &\rightarrow \Phi(A_3), \\ A_4 &\rightarrow \Phi(A_4), y \rightarrow \Phi(y) \end{aligned} \quad (21)$$

In this paper, some non-linear data are retained to feature space by multi-kernel function mapping,

$$K(A_1, A_1), K(A_2, A_2), K(A_3, A_3), K(A_4, A_4)$$

$$K(A_1, y), K(A_2, y), K(A_3, y), K(A_4, y) \text{ are obtained.}$$

4) Dictionary learning without occlusion

$\Phi(A_1), \Phi(A_2), \Phi(A_3), \Phi(A_4)$  are trained respectively to the dictionary  $D_1, D_2, D_3, D_4$

$$D_i = \arg \min \{ \|\Phi(y) - \Phi(A_i) \alpha\|_2^2 \} \quad (22)$$

Actually, Equation (22) cannot be solved directly, which can be rewritten as:

$$D_i = \arg \min \{ \|\Phi(A_i)^T (\Phi(y) - \Phi(A_i) \alpha)\|_2^2 \} \quad (23)$$

According to the definition of kernel function

$$k(x, y) = \langle \Phi(x), \Phi(y) \rangle = \Phi(x)^T \Phi(y) \quad (24)$$

The kernel function in (23) can also be written as follows:

$$\begin{aligned} \Phi(A_i) &= [\Phi(v_{i,1}), \Phi(v_{i,2}), \dots, \Phi(v_{i,m_i})] \\ B &= \begin{bmatrix} k(v_{i,1}, v_{i,1}) & k(v_{i,1}, v_{i,2}) & \dots & k(v_{i,1}, v_{i,m_i}) \\ k(v_{i,2}, v_{i,1}) & k(v_{i,2}, v_{i,2}) & \dots & k(v_{i,2}, v_{i,m_i}) \\ \dots & \dots & \dots & \dots \\ k(v_{i,m_i}, v_{i,1}) & k(v_{i,m_i}, v_{i,2}) & \dots & k(v_{i,m_i}, v_{i,m_i}) \end{bmatrix} \\ & \quad \text{M} \quad \text{M} \quad \text{O} \quad \text{M} \\ \Phi(A_i)^T \Phi(A_i) &= B \end{aligned} \quad (25)$$

$$\begin{aligned} \Phi(A_i)^T \Phi(y) &= [\Phi(v_{i,1}), \Phi(v_{i,2}), \dots, \Phi(v_{i,m_i})]^T \Phi(y) \\ &= [k(v_{i,1}, y), k(v_{i,2}, y), \dots, k(v_{i,m_i}, y)]^T \end{aligned} \quad (26)$$

Solving Equation (26),  $D_1, D_2, D_3, D_4$  are obtained.

5) Dictionary learning with occlusion: In the case of occlusion Equation (18) can be written:

$$\Phi(y) = D \alpha_0 + e_0 \quad (27)$$

where,  $e_0$  is occlusion, the above (27) is written as:

$$\Phi(y) = [D, I] \begin{bmatrix} \alpha_0 \\ e_0 \end{bmatrix} = K W_0 \quad (28)$$

where,  $I$  is unit matrix. Then  $D_1, D_2, D_3, D_4$  is converted to  $K_1, K_2, K_3, K_4$  then to solve the minimum norm of  $l_{2,1}$

$$\begin{aligned} \hat{W} &= \arg \min \{ \|\Phi(y) - K_i W\|_2 + \lambda \|W\|_{2,1} \} \\ s.t. & \lambda > 0 \end{aligned} \quad (29)$$

The variety of the sample with the smallest error is the maize variety.

$$y' = \arg \min \{ \|\Phi(y) - K_i W\|_2 \} \quad (30)$$

To ensure the consistency of train samples and test samples, the same feature matrix and combined multi-kernel matrix are used in the process of reducing dimensional and adding kernel function.

## 4 Experiment and discussion

In the experiment, four types of maize varieties Liangyu 99, Denghai 6702, Delinong 988 and Zhong 909 are selected, and 200 samples for each type variety are chosen randomly, in which 160 samples are selected as training samples. The corn kernels images are collected by camera shooting, which are preprocessed firstly to

benefit the feature extracting. The picture is with BMP format and the pixel is 521×341.

**4.1 Image preprocessing**

The purpose of image preprocessing is to improve image data construction for valuable information extracting and useless information weakening. In this paper, the grain images are preprocessed by median filtering, image segmentation and morphological analysis.

The gray scale transformation enhances the gray scale range and enriches the gray level, results in that the image becomes more clear and the feature becomes more obvious; The median filter provides smooth processing, which changes the edges of the image more clear; The threshold segmentation is adopted to affirm the landmark characteristic region; In morphological operation, the open operation is introduced to smooth the contours of an image and eliminate the interference of fine objects, by which the objects are separated at the slim points and the boundaries of the larger objects are smoothed without changing their size significantly.

The pretreating result is shown in Figure 3.

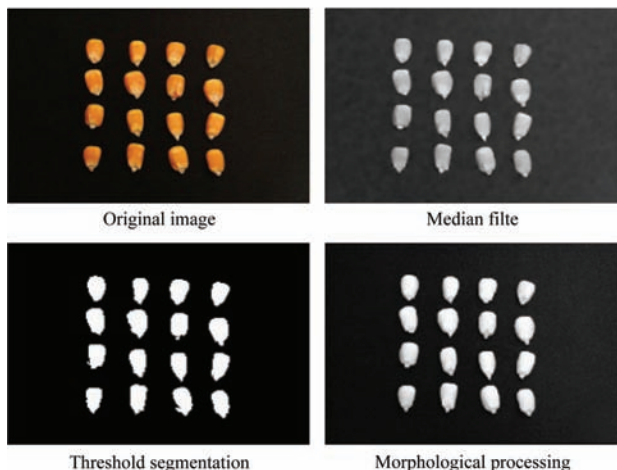


Figure 3 Result of image preprocess

**4.2 Maize grain classification**

For a more intuitive assessment of identification performance, the recognition rate (RR) is adopted:

$$RR = \frac{\text{the number of correctly classified}}{\text{the total in umber of tests}}$$

For training samples, the appearance characteristic parameters are extracted as the typical feature of corn kernels to distinguish different types of corn, which consists a classifying dictionary of corn kernels and is shown in Figure 4.

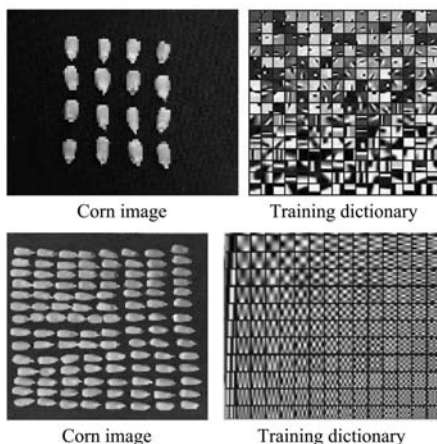


Figure 4 Feature dictionary

**4.2.1 Complete maize classification**

For complete kernel classification, 10% data of each corn variety sample set is selected randomly as a test sample, and the others is used as a training sample. The single hidden layer structure with only one hidden layer in the middle layer is used in BP neural network model. In order to ensure the accuracy and effectiveness of the experiment, 30 repeated experiments are carried out for each set of data, taking the average value. The experiment results of the approach in this paper are compared with the experiment results of traditional classification such as the SRC method in [27], the BP method in [28], the SVM method in [29] and KNN method, which is shown in Table 1.

**Table 1 Selection of recognition ratios for different proportions of test samples**

	10%	20%	30%	40%
SRC	0.92	0.91	0.88	0.86
BP	0.81	0.80	0.85	0.93
SVM	0.82	0.83	0.87	0.84
KNN	0.81	0.835	0.852	0.89
Our Algorithm in this paper	0.92	0.92	0.94	0.94

The experimental results show that the recognition rate of the method proposed in this paper is higher than the other four methods mentioned in the paper under the same data set. The method proposed in this paper also has the following advantages: 1) This method shortens the training time by reducing dimension and K-SVD dictionary. It solves the problems of BP neural network and KNN classifier, which have a long training and recognition time. 2) The multi-core method which is introduced in this paper will reduce the characteristic loss of nonlinear data and improve the recognition rate. This solves the problem of low recognition rate caused by the support vector machine approach which cannot completely describe the characteristics of the maize sample (Table 1). 3) When the proportion of the test samples is higher than 30%, compared with the widely used SRC method, the average recognition rate of the method proposed in this paper is higher; 6% than the SRC method, In order to show the superiority of the approach proposed in this paper more clearly the data in Table 1 is generated as a chart which is shown in Figure 5.

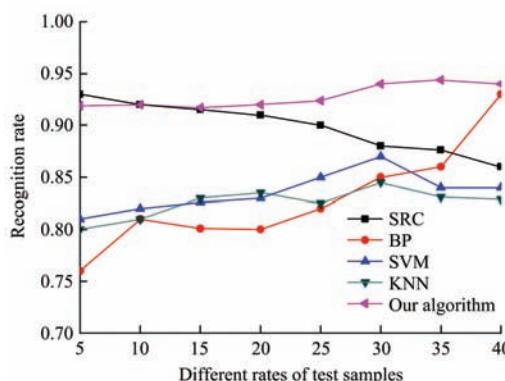


Figure 5 Comparison of recognition rates under different proportions of samples

It is shown in Figure 5 that the recognition rate of the method proposed in this paper increases with the size of the test sample proportion. The recognition effect of SVM and SRC is far behind the method proposed in this paper. The recognition rate of BP network does rise with the increase of test sample proportion to be almost equal to the method proposed in this paper. However, several parameters in the BP network and SVM need to be

optimized some times in order to get a higher recognition rate in the experiment. The running time of the method proposed in this paper is about half of BP network and 1/3 of SVM. The choice of parameters of the method proposed in this paper is relatively simple, only the minimum error and the number of iterations are considered, and by these the higher classification rate can be obtained. The method proposed in this paper has faster computing speed and better robustness comparing with BP network and SVM algorithm.

#### 4.2.2 Incomplete maize classification

For incomplete kernel classification, 20% data of each corn variety sample set is selected randomly as the test sample, and the rest is used as training samples. The occlusion rate is 10% to 20% and the occlusion position randomly. Part of incomplete kernel is shown in Figure 6. The occlusion-sparse-representation-classifier is adopted for incomplete grain classification. The experiment carries out 10 times for each dataset, taking the mean value. The comparison of the experiment results of approach used in this paper with the experiment results of SRC, BP, KNN and SVM are shown in Table 2.



Figure 6 Schematic diagram of the partially blocked grain

At the same time, 10% of the whole dataset was randomly selected as the test samples, and the rest as training samples. The difference is that the contrast method which SRC is replaced by the sparse representation classifier with occlusion. For different data sets, different methods and different occlusion, the experiment can be different recognition rate as shown in Table 2 (In the experiment, each set of data was made 10 times, taking the mean value)

**Table 2 Recognition rates of different shading rates**

	10%	20%	30%	40%
SRC	0.94	0.90	0.86	0.80
BP	0.93	0.89	0.83	0.81
SVM	0.91	0.86	0.81	0.79
KNN	0.915	0.875	0.845	0.817
Our Algorithm	0.96	0.94	0.89	0.83

The experiment results analysis. 1) The recognition rate of each method is decreased with the increasing of the shading rate. 2) By contrast, the recognition rate of the method used in this paper is higher than that of SRC, BP, KNN and SVM. For the reason that SRC in the sparse representation is used to face recognition and the SVM method put forward without considering the occlusion problem.

The data in Table 2 are generated as Figure 7 to show the superiority of the method set out in this paper more clearly.

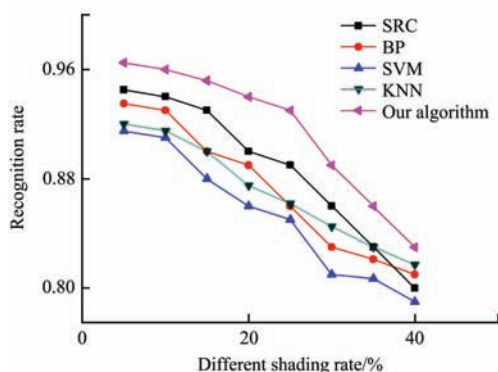


Figure 7 Recognition rate of different shading rate

Figure 7 shows the different recognition rates in the data set of the 5 approaches under different occlusion levels. The transverse coordinate represents different amounts from 5% to 40% and the longitudinal axis represents the recognition rate. The results show that the recognition effect of the method proposed in this paper is demonstrably better than the other four methods.

## 5 Conclusions

In this study, aiming at the maize varieties classification and identification under complex background and partial occlusion, a multi-kernel based dictionary training method was put forward, which was combined with sparse representing to identify and classify corn images. In order to avoid the huge amount of calculation from taking all the training images as the calculation object of sparse representation, the *K*-SVD dictionary training method was adopted, by which part of the "optimal base" was selected to represent all training images. Considering the problem that non-linear data could not be processed in general low-dimensional space, the kernel method was introduced. Using kernel method, the different characteristics of the picture were fully considered, and the weight for each characteristic was distributed automatically, from which a high classification accuracy was obtained, and the calculating speed was also improved.

## Acknowledgments

We acknowledge that this work was financially supported by the National Natural Science Foundation of China (Grant No. 61472172, 61673200 and 61471185) and by the Natural Science Foundation of Shandong Province of China (Grant No. ZR2016FM15).

## [References]

- [1] Liu C C, Shaw J-T, Poong K-Y. Classifying paddy rice by morphological and color features using machine vision. *Cereal Chemistry*, 2005; 82(6): 649–653.
- [2] Jinrose M, Prachayawarakorn S, Soponronnarit S, Devahastinet S. Development of a computer vision system and novel evaluation criteria to characterize color and appearance of rice. *Drying Technology*, 2010; 28(9): 1118–1124.
- [3] Yang W, Winter P, Sokhansanj S, Wood H, Crerer B. Discrimination of Hard-to-pop Popcorn Kernels by Machine Vision and Neural Networks. *Biosystems Engineering*, 2005; 91(1): 1–8.
- [4] Kurtulmuş F, AliBaş I, Kavdır I. Classification of pepper seeds using machine vision based on neural network. *Int J Agric & Biol Eng*, 2016; 9(1): 51–62.
- [5] Guo D, Zhu Q, Huang M, Guo Y, Qin J. Model updating for the classification of different varieties of maize seeds from different years by hyperspectral imaging coupled with a pre-labeling method. *Computers & Electronics in Agriculture*, 2017; 142: 1–8.
- [6] Zheng Y, Zhu Q, Huang M, Guo Y, Qin J. Maize and weed classification using color indices with support vector data description in outdoor fields. *Computers & Electronics in Agriculture*, 2017; 141: 215–222.
- [7] Ambrose A, Kandpal L M, Kim M S, Lee W H, Cho B K. High speed measurement of corn seed viability using hyperspectral imaging. *Infrared Physics & Technology*, 2016, 75: 173–179.
- [8] Vithu P, Moses J A. Machine Vision System for Food Grain Quality Evaluation: A Review. *Trends in Food Science & Technology*, 2016, 56: 13–20.
- [9] Xie C, He Y. Modeling for mung bean variety classification using visible and near-infrared hyperspectral imaging. *International Journal of Agricultural & Biological Engineering*, 2018, 11(1): 187–191.
- [10] Hao S, Wang W, Yan Y, Bruzzone L. Class-wise dictionary learning for hyperspectral image classification. *Neurocomputing*, 2016, 220.
- [11] Zhang Y, Xu T, Ma J. Image Categorization using Non-negative Kernel Sparse Representation. *Neurocomputing*, 2017.
- [12] Yang S Y, Han Y, Zhang X R. A sparse kernel representation method for

- image classification. International Joint Conference on Neural Networks. IEEE, 2012; 1–7.
- [13] Zhang L, Zhou W D, Chang P C, Liu J, Yan Z, Wang T. Kernel Sparse Representation Based Classifier. IEEE Trans on Signal Processing, 2012, 60(4): 1684–1695.
- [14] Li H, Gao Y, Sun J. Fast Kernel Sparse Representation. International Conference on Digital Image Computing Techniques and Applications. IEEE, 2011: 72–77.
- [15] Meng J, Jung C. Class-discriminative kernel sparse representation-based classification using multi-objective optimization. IEEE Transactions on Signal Processing, 2013; 61(18): 4416–4427.
- [16] Gao S, Tsang I W, Chia L T. Sparse representation with kernels. IEEE Transactions on Image Processing. Publication of the IEEE Signal Processing Society, 2013; 22(2): 423–434.
- [17] Nguyen H V, Patel V M, Nasrabadi N M, Chellappa R. Kernel dictionary learning. IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2012; pp.2021–2024.
- [18] Van N H, Patel V M, Nasrabadi N M, Chellappa R. Design of non-linear kernel dictionaries for object recognition. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2013; 22(12): 5123–5135.
- [19] Golts A, Elad M. Linearized kernel dictionary learning. IEEE Journal of Selected Topics in Signal Processing, 2015; 10(4): 726–739.
- [20] Zare T, Sadeghi M T. A Novel multiple kernel-based dictionary learning for distributive and collective sparse representation based classifiers. Neurocomputing, 2017; 234: 164–173.
- [21] Xu Z, Jin R, King I, Lyu M R. An extended level method for efficient multiple kernel learning. In Advances in Neural Information Processing Systems, 2009; pp.1825–1832.
- [22] Rakotomamonjy A, Bach F R, Canu S, Grandvalet Y. Simplemkl. Journal of Machine Learning Research, 2008; 9(3): 2491–2521.
- [23] Xu Z, Jin R, Yang H, Lyu M R. Simple and efficient multiple kernel learning by group lasso. International Conference on Machine Learning. DBLP, 2010; pp.1175–1182.
- [24] Wen X Z, Fang W, Zheng Y H. An algorithm based on Haar-like features and improved AdaBoost classifier for vehicle recognition. Acta Electronica Sinica, 2011; 39(5): 1121–1126.
- [25] Kim S J, Koh K, Lustig M. An interior-point method for large-scale  $L_{2,1}$ -regularized least squares. IEEE Journal on Selected Topics in Signal Processing, 2007; 1(4): 606–617.
- [26] Bosch A, Zisserman A, Munoz X. Representing shape with a spatial pyramid kernel. ACM International Conference on Image and Video Retrieval. ACM, 2007: 401–408.
- [27] Yang J, Tian Y, Duan L Y, Huang T, Gao W. Group-sensitive multiple kernel learning for object recognition. IEEE Transactions on Image Processing, 2012; 21(5): 2838–2852.
- [28] Feng L J, Li X J, Wen C L. Wheat varieties identification research based on sparse representation. Journal of Jiangnan University: Natural Science Edition, 2015; 14 (6): 730–735.
- [29] Yang S Q, Ning J F, He D J. Identification of varieties of rice based on sparse representation. Transactions of the CSAE, 2011; 27(3): 191–195.