ISSN : 0974 - 7435 Volume 10 Issue 5

BioTechnology
An Indian Journal

FULL PAPER

BTAIJ, 10(5), 2014 [1013-1020]

# Detection of hidden insect of wheat by biological photon technique

**Shi Weiya*, Qiao Nana, Liang Yitao, Wang Feng**
**School of Information Science and Engineering, Henan University of Technology, Henan Zhengzhou 450001, (CHINA)**
**E-mail : swymail@126.com**

## ABSTRACT

In order to prevent grain mass and quality loss, a fast and efficient method for early detection of insect infestation of grain is urgently needed during trade and storage. Based on the biophoton analytical technology (BPAT), this work adopted a new method of extracting feature by combining statistical characteristics and histogram distribution. Considering the sample covariance matrices of any single class could be singular, the feature vector was compressed by principal component analysis (PCA) and given as inputs to classifiers for the identification of uninfested wheat and infested wheat, such as linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), mahalanobis and linear support vector machines (linear SVM). For further improving the classification accuracy, regularized discriminant analysis (RDA) was presented to optimize QDA and mahalanobis algorithms. The results proved that the proposed method is workable.


## KEYWORDS

Biophoton;
Statistical characteristics;
Histogram features;
Identification.

## INTRODUCTION

According to the FAO's survey, the losses caused by stored grain pests in 50 countries reach the total yield of 5% to 10%, even as high as 30%. One of the most serious pests endangering the safety of stored grain in China is *Sitophilus zeamais*. The grain damage rate of *Sitophilus zeamais* is ~25.7%, the actual loss rate is up to 5.65%[1,2]. In order to control pests accurately in the time, a nice early detection method of hidden insect in bulk grain can help to reduce the loss of stored grain. At present, the detection methods include traditional detection method[3], chemical detection method[4] and new nondestructive testing method such as microwave radar method[5], acoustic method[6] and the near infrared spectroscopy[7]. Traditional methods are complicated operation and poor accuracy; Chemical detection methods do not belong to the green nondestructive testing, which is not guaranteed real time and difficult to control costs due to chemical substances involved. Existing non-destructive testing methods in early growing stage (especially egg stage) is inefficient or relatively lagged. The presentation of BPAT provides a novel idea in the detection of hidden insects. Compared with the other methods, BPAT can offer the whole information on the organism changes caused by various internal stations or environmental impacts.

Biological ultra weak luminescence (UWL) is a common phenomenon in nature, existing in a variety of animals, plants and microorganisms[8]. Biological ultra weak luminescence can be broadly classified into two types: spontaneous luminescence (SL) and delayed luminescence[9].

The study focused on designing classifiers. Wheat kernels and *Sitophilus zeamais* were used as research objects. Based on the self-illumination characteristics of uninfested wheat and infested wheat, grain characteristic vector (GCV) was constructed by analyzing extracted the histogram feature[10] and statistical characteristic values such as mean, variance. Then GCV was put into classifiers (LDA[11], QDA[12], Mahalanobis, Linear SVM). The proposed model was non-destructive and accurate detection methods, which provides scientific and reliable basis for management decisions of stored grain insects.

## EXPERIMENTAL MODEL

### Sample testing

Wheat kernels (*Zheng Mai 7698*) harvested in 2013 were used. The kernels were washed three times by distilled water and placed in the drying oven to dry until the moisture content of kernels was up to (13.7±0.2) %. Then they were kept in a thermostatic box at a temperature of 28°C.

### (a) Sample preparation

Part of wheat kernels were taken out from the thermostatic box for infectious treatment. 20 adults (*Sitophilus zeamais*) and 20g kernels were put in a small airtight container (28°C). Two days later, the adults were picked out and the kernels contained *Sitophilus zeamais* of egg stage, which were chosen as infested samples.

### (b) Testing

Before UML was measured, the BPCL should be open. High voltage was regulated to 1030V and measurement chamber was controlled at 28°C (the optimum temperature for *Sitophilus zeamais* growth). The BPCL should be preheated 30 min in order to make the bottom reach steady state.

Kernels were weighed with accuracy to a mass of one grain, so that they obtained a mean mass of 5.602±0.023g. To avoid recording the light-induced luminescence from prepared samples of kernels, they were kept for 30min in complete darkness prior to the start of UML detection. Laboratory apparatus is shown in Figure 1. The measurement time was 1800s and the interval was set to 1s in measuring software (Figure 1.1). In order to avoid the impact of external factors on the measurements of UML, measurement chamber (Figure 1.2) and signal analyzer host (Figure 1.3) should be kept in complete darkness, and laboratory temperature was controlled at (20±1) °C.

The background noise would be automatically subtracted by the instrument after the measurement of UML. Due to the counted random data from 201s to 1224s with stability, each sample would select spontaneous photon counts between 201 s to 1224 s. The spontaneous biophoton emission curve of wheat kernels in Figure 2 fluctuates significantly—


**1. Measuring software, 2. Measuring chamber, 3. Signal analyzer host, 4. Data processing system**

**Figure 1 : Laboratory apparatus**

besides the influence of machine noise, it is speculated that the interaction between internal metabolism of wheat kernels produces different photon emission signals.

## Feature extraction

### (a) Histogram feature extraction

When data of samples were described using frequency distribution histograms[13], appropriate grouping and statistic would make data distribution more clear. By analysis, it was proper that the date of each sample was divided into thirteen groups if the magnitude of class interval was 5. Owing to the negative in photons counts after subtracting the background was classified as zero, it would not be counted in first group. The thirteen values were selected as part of the eigenvalues. In Figure 3, the frequency of uninfested wheat falling in the region of 5, 10, 15, 20 is relatively more than uninfested wheat. However, the spontaneous photon counts of uninfested wheat falling in the region of 40, 45, 50, 55, 60, 65 were almost zero, and photon counts of infested wheat were relatively large.

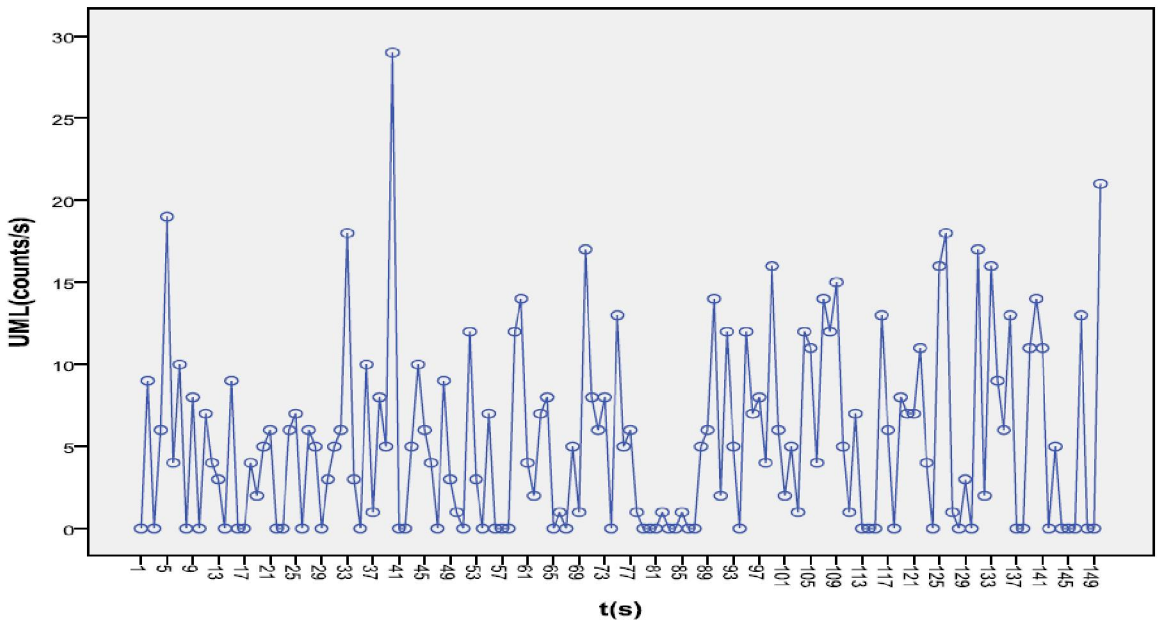


**Figure 2 : The spontaneous biophoton emission curve of wheat kernels—the abscissa is time and the vertical coordinate is spontaneous photon counts**

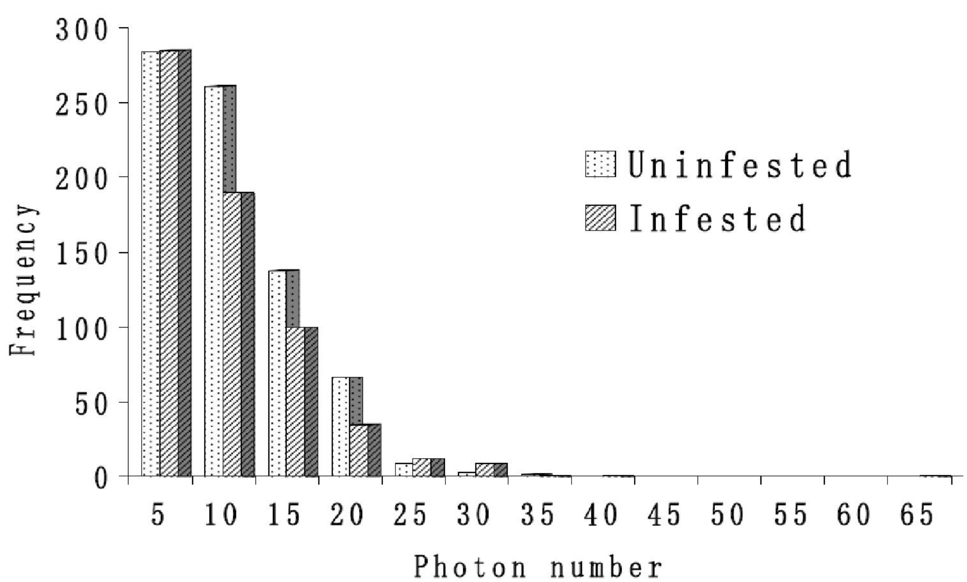


**Figure 3 : The histogram distribution of uninfested wheat and infested wheat**

### (b) Statistical feature extraction

The statistical characteristic values of each sample were calculated respectively, including mean, variance, median, quartile, mean difference, discrete coefficient, skewness, kurtosis, a total of eight characteristics.

### (c) Composition grain feature vector

Eight statistical characteristics and thirteen histogram features were combined to form GCV as the import of the classifiers.

## Modeling

Considering that small sample could cause the singularity and instability of covariance matrices of any single class, the PCA was performed for GCV[14,15]. Then four classsifiers (LDA, QDA, Mahalanobis, linear SVM) were adopted to identify uninfested wheat and infested wheat. Meanwhile, RDA[16] was presented to improve the classification accuracy of QDA and Mahalanobis (the GCV without compressed by PCA). The optimal classifier was selected by comparing their recognition rates.

### (a) Data dimension reduction

PCA is a multivariate statistical method which investigates correlation between multiple variables, to study how to reveal the internal structure between multiple variables by a few principal components. That is, a few principal components derived from the original variables are made to keep the information of original variables as much as possible and unrelated to one another.

### (b) Classification algorithms

The tasks of discriminant analysis are to minimize misjudgment objects by establishing a better discriminant function according to the available samples classified clear—for a given sample, the classifier can judge which overall it came from.

### (1) LDA

LDA mainly makes high-dimensional patterns of samples project the best identification vector space to achieve the effect of extracting classification information and compressing feature space dimension. After the projection, model sample has the largest distance between classes and the smallest distance within the class in the new subspaces. The discriminant function of LDA is shown at Eq. (1). It selects Φ which makes J (Φ) reach maximum as projection direction.

$$J(\Phi)=\frac{\Phi^T S_b \Phi}{\Phi^T S_w \Phi} \quad (1)$$

$$S_b=\sum_{i=1}^{c} n_i (u_i-u)(u_i-u)^T \quad (2)$$

$$S_w=\sum_{i=1}^{c}\sum_{x_k\in class\,i}(u_i-x_k)(u_i-x_k)^T \quad (3)$$

Φ is n column vector and $S_b$ is between-class scatter matrix (Eq. (2)). $S_w$ is within-class scatter matrix (Eq. (3)). The $u_i$ is sample mean of the class i and u is the overall sample mean. The $n_i$ shows the number of training samples in class i, c shows the total number of sample classes and $x_i$ denotes the sample i.

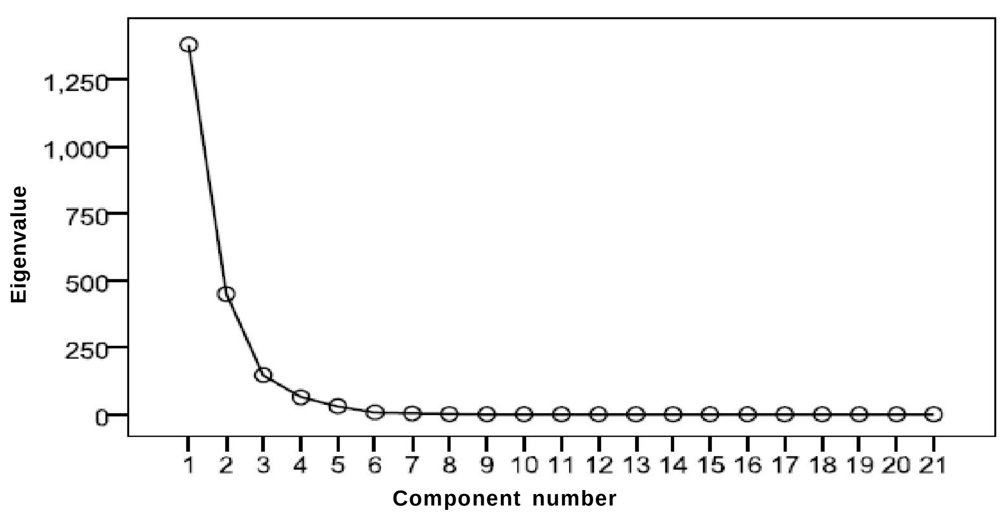


**Figure 4 : Characteristic value curve of 21 compositions**

**(2) QDA**

Based on have mastered sample information for any single class, the Bayesian discriminant principle is to establish discriminant function by suming up the objective regularity of classification. It is divided into two classes (LDA and QDA) according to the different orders of discriminant function. In the Bayesian decision theory, the sample is attributed to classes with the largest posterior probability under classes and their probability distribution have been known in advance. QDA assumes that prior probabilities of any classes are the same and conditional probability density adopts the normal distribution.

$$g_i(x) = -\frac{1}{2}\log(|\sum_i|) - \frac{1}{2}(x-u_i)^T \sum_i^{-1}(x-u_i) \quad (4)$$

$$\sum_i = \frac{1}{n_i}\sum_{j=1}^{n_i}(x_{ij}-u_i)(x_{ij}-u_i)^T \quad (5)$$

The QDA is shown at Eq. (4). The two variables, $u_i$ and $\Sigma_i$, are belong to the maximum likelihood estimation. Covariance matrix of class i ($\Sigma_i$) is indicated at Eq. (5). $x_{ij}$ denotes training sample j of class i. Assuming that the total samples can be divided into two classes, for any sample x, if $g_i(x) > g_j(x)$, x can be attributed to the former.

**(3) Mahalanobis**

Mahalanobis, a covariance distance, is an effective method to calculate the similarity of two unknown sample sets. In the Eq. (6), X and $G_i$ denote a sample set and overall of any single class respectively.

$$D(X, G_i) = \sqrt{(X-u_i)^T \Sigma_i^{-1}(X-u_i)} \quad (6)$$

For the recognition of the two classes, overall G1 and G2, a given sample X is judged by the discriminant rule : if $D^2(X, G_1) \le D^2(X, G_2)$, X is determined to $G_1$, or it will be attributed to $G_2$.

**(4) Linear SVM**

SVM makes the input vector map to a high dimensional feature space using a kernel function, then constructs an optimal hyperplane to approximate classification function in the space. The SVM selects C– SVC. The loss parameter is set to 5 and kernel function selects linear kernel function. The experimental data should be normalized before processing.

**(5) RDA**

RDA was mainly used to settle the singularity and instability of class covariance matrix in the calculation of QDA and Mahalanobis. It contains complex parameter $\lambda$ and contraction parameter $\gamma$.

$$\sum_i^{\lambda} = \frac{(1-\lambda)S_i + \lambda S}{(1-\lambda)n_i + \lambda n} \quad (7)$$

$$\sum_i^{\lambda,r} = (1-\lambda)\sum_i^{\lambda} + \gamma c_i(\lambda) I_p \quad (8)$$

$$S_i = n_i \sum_i, \quad S = \sum_{i=1}^{c} n_i \Sigma_i$$

In the class covariance matrix (Eq.(7)), Parameter $\gamma$ is adopted to further adjust Eq. (8). P reveals the dimensions of the pattern feature subspace and $I_p$ is the unit matrix (P * P). The values of $\lambda$ and $\gamma$ range from 0 to 1.

## RESULTS AND DISCUSSION

To verify the validity of these classifiers, selecting uninfested wheat and infested wheat (containing insects for 24-28 days) as the experimental samples, including the training samples (thirty groups) and the test samples (twenty groups). In this work, algorithms were written using MATLAB software. Linear SVM was designed by means of MATLAB toolbox function, mainly including training function *svmtrain* and forecast function *svmpredict*.

### The results and analysis

When PCA was applied to the GCV, the first four principal components were selected. The inclination between composition 3 and 4 is larger than component 4 and 5, and the component 4 can be regarded as the turning point of the curve (Figure 4). By comparing the cumulative contribution rate of component 4 (97%) and component 5 (99%), the component 5 cannot provide more information. So it was appropriate to compress the feature to four dimensions.

Four classifiers (LDA, QDA, Mahalanobis, linear SVM) were applied to train and test for samples. Their classification accuracies were presented in TABLE 1. It was observed that LDA and linear SVM identified about 90% of kernels infested, which were higher than classification accuracies of QDA (75%) and Mahalanobis (80%). More intuitive classification histogram was displayed in Figure 5. The LDA and linear SVM classifiers accurately discriminated 90% of uninfested and 90% of infested kernels. Whereas, QDA gave a classification accuracy of 80% and 70% for uninfested and infested kernels, Mahalanobis classifiers identified 90% of uninfested and 70% of infested kernels. Comprehensive comparison reveals that LDA and linear SVM classifier is better than QDA and Mahalanobis, and Mahalanobis recognition rate is higher than QDA.

Results showed that the classification effects of QDA and Mahalanobis were relatively low. For the purpose of resolving the singularity and improving identification accuracy of QDA and Mahalanobis, the class covariance matrix of them could be treated by RDA instead of PCA. With the change of parameter values ($\lambda$ and $\gamma$), the final classification results of RDA were different. The classification accuracies of RDA were demonstrated at TABLE 2. When the complex parameter $\lambda$ and contraction parameter $\gamma$ were appropriate, the optimal classification results of QDA and Mahalanobis were up to 90%.

## Discussion

In practice, considering the consequences caused by different classification results, mainly including that infested wheat are mistaken for uninfested wheat, miss alarm and false alarm should be taken into consideration[17]. Miss alarm reveals how many infested wheat were sentenced to uninfested wheat and false alarm demonstrates how many groups were uninfested wheat in sentenced infested wheat. After the GCV were compassed by PCA, miss alarm of LDA and linear SVM (10%) were much lower than QDA and Mahalanobis (30%), similar to false alarm rate (TABLE 3). While false alarm of QDA (18.2%) were slightly higher than Mahalanobis (12.5%), and the latter classification accuracy was surpass the former in TABLE 1.

When the singularity of class covariance matrix was settled by RDA, the classification accuracy was improved with the highest classification accuracy (90%) using QDA and Mahalanobis classifiers, respectively using the combined features in the model (TABLE 2). In TABLE 2 and TABLE 4, QDA gave a classification accuracy of 90% with the parameter values ($\lambda$, $\gamma$) at 0.1–0.5, the miss alarm rate of 0% and false alarm rate of 16.7%. When parameter values ($\lambda$, $\gamma$) were 0.8, the classification accuracy of QDA was also 90%, but miss alarm rate and false alarm rate were 10%. For Mahalanobis classifier, with the parameter values ($\lambda$, $\gamma$) ranging from 0.5 to 0.7, the kernels infested could be discriminated with the highest classification accuracies

**TABLE 1 : The classification accuracies of uninfested wheat and infested wheat using eigenvector compressed by PCA**

| Classifiers | Correct (group) | Failure (group) | Accuracy(%) |
|---|---|---|---|
| LDA | 18 | 2 | 90 |
| QDA | 15 | 5 | 75 |
| Mahalanobis | 16 | 4 | 80 |
| Linear SVM | 18 | 2 | 90 |

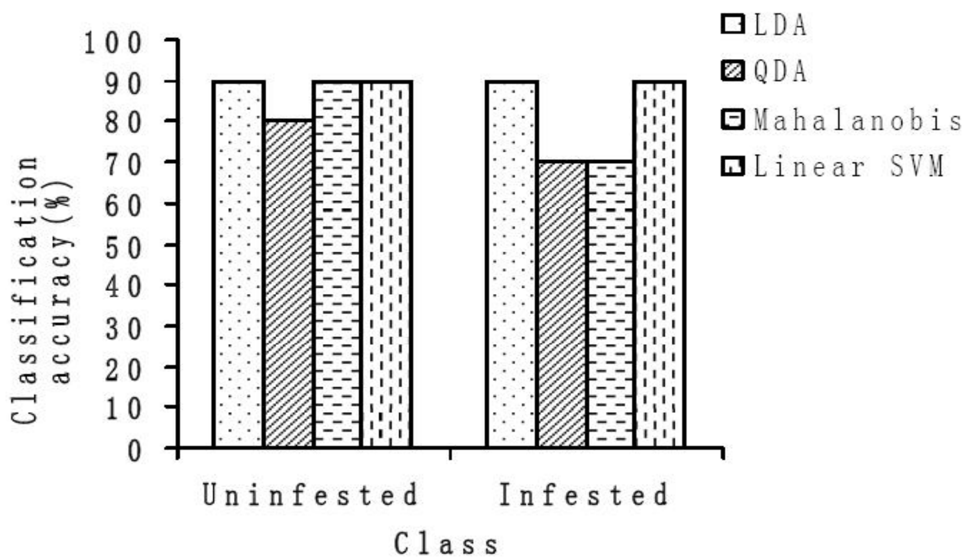


**Figure 5 : Two-way classification accuracies of uninfested wheat and infested wheat using eigenvector compressed by PCA**

**TABLE 2 : The classification accuracies of QDA and Mahalanobis improved by RDA**

| λ,γ | Class covariance matrix treated by RDA (%) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
| QDA | 90 | 90 | 90 | 90 | 90 | 85 | 84 | 90 | 85 |
| Mahalanobis | 85 | 85 | 85 | 85 | 90 | 85 | 85 | 90 | 85 |

**TABLE 3 : The miss alarm and false alarm of four classification algorithms**

| Statistics (%) | eigenvector compressed by PCA | | | |
|---|---|---|---|---|
| | LDA | QDA | Mahalanobis | Linear SVM |
| Miss alarm | 10 | 30 | 30 | 10 |
| False alarm | 10 | 18.2 | 12.5 | 10 |

**TABLE 4 : The miss alarm and false alarm of QDA and Mahalanobis improved by RDA**

| Statistics (%) | Class covariance matrix treated by RDA (λ,γ) | | | | | | |
|---|---|---|---|---|---|---|---|
| | QDA | | | | Mahalanobis | | |
| | 0.1-0.5 | 0.6-0.7 | 0.8 | 0.9 | 0.1-0.4 | 0.5-0.7 | 0.8-0.9 |
| Miss alarm (%) | 0 | 10 | 10 | 20 | 20 | 10 | 20 |
| False alarm (%) | 16.7 | 18.2 | 10 | 11.1 | 11.1 | 10 | 11.1 |

of 90%, miss alarm rate and false alarm rate were 10%. Obviously, the classification results of QDA and Mahalanobis improved by RDA were higher than treated by PCA. By the comparison of four classifiers and considering relatively high demand for miss alarm, QDA classifier was found with the highest classification accuracy in discriminating the kernels infested when QDA was optimized by RDA with parameter values ($\lambda$, $\gamma$) at 0.1–0.5. Meanwhile, the classification result of Mahalanobis handled by RDA ($\lambda$, $\gamma$ at 0.5–0.7), was comparable to the effects of LDA and linear SVM classifiers treated by PCA.

## CONCLUSION

In the study, spontaneous light photon counts of uninfested wheat and infested wheat was measured by instrument (BPCL). Furthermore, the statistical feature selection and histogram feature extraction are completed. The test results indicated that the four classifiers could automatically determine the kernels infested after these eight statistical features (mean, variance, median, quartiles, mean difference, dispersion coefficient, skewness, kurtosis) with thirteen histogram features were combined. And the kernels infested were identified to be more significant using QDA with RDA.

The results also show that the proposed model can distinguish uninfested and infested wheat better. It benefits for the early detection of grain insects, which provides a scientific basis for taking appropriate action as possible. However, the classifiers also can not detect *Sitophilus zeamais* of egg stage and larval stage, which needs further research and exploration.

## ACKNOWLEDGEMENTS

This work is sponsored by National Natural Science Foundation of China (31171775, 11201120 & 61174056), National High Technology Research and Development Program of China (2012AA101608), Province Technology R & D Program of Henan Province of China (142102210148, 112102210190), Natural Science Foundation of Henan Educational Committee (14B550002) and Developing issue of Zhengzhou Technology Division (2010SFXM470). Research project was supported by the Key Laboratory of Grain Information Processing and Control.

## REFERENCES

**[1]** LS/T 1709-2004. Classification and codes of grain information - Classification and codes of pest and disease in stored-grain[S], Beijing, Standards Press of China, **(2004)**.

**[2]** Ma Xiaohui, Wang Dianxuan, Li Keqiang, Guo Chao; Investigation on stored grain insect species and resistance to phosphine in state grain reserves of China[J], GRAIN Storage, **37(1)**, 7-10 **(2008)**.

**[3]** Bai Xuguang; A review of technology for detecting stored grain in sects[J], GRAIN Storage, **39(1)**, 6-9 **(2010)**.

**[4]** Chen Zhaopeng; The hidden and non hidden in Stored Grain Pests Detection Technology[J], GRAIN Storage, **24(5-6)**, 96-100 **(1995)**.

**[5]** R.W.Mankin; Microwave radar detection of stored-product insects[J], Journal of Entomology, **97(3)**, 1168-1173 **(2004)**.

**[6]** D.W.Hagstrum, P.W.Flinn, D.Shuman; Automated monitoring using acoustical sensors for insects in farm-stored wheat[J], Journal of Economic Entomology, **89(1)**, 211-217 **(1996)**.

**[7]** E.B.Maghirang, et al.; Automated detection of single wheat kernels containing live or dead insects using near-infrared reflectance spectroscopy[J], American Society of Agricultural Engineers, ISSN 0001-2351, **46(4)**, 1277-1282 **(2003)**.

**[8]** Gu Qiao; Biophotonics[M], Beijing, Science Press, **(2007)**.

**[9]** Li Shuli; Study on Ultra Weak Luminescence in Tomato Ripening Process[D], Northwestern University, Master's Thesis, **(2005)**.

**[10]** S.Kaliramesh, V.Chelladurai, D.S.Jayas, K.Alagusundaram, N.D.G.White, P.G.Fields; Detection of infestation by callosobruchus maculatus in mung bean using near-infrared hyperspectral imaging[J], Journal of Stored Products Research, **53**, 107-111 **(2013)**.

**[11]** W.Wu, Y.Mallet, B.Walczak, W.Penninckx, D.L.Massart, S.Heuerdingb, F.Ernib; Comparison of regularized discriminant analysis, linear discriminant analysis and quadratic discriminant analysis, applied to NIR data[J], Analytica Chimica Acta, **329**, 257-267 **(1996)**.

**[12]** Wang Wei-Dong, Yang Jing-Yu; Quadratic discriminant analysis method based on virtual training samples[J], Acta Automatica Sinica, **34(4)**, 401-407 **(2008)**.

**[13]** C.B.Singh, D.S.Jayas, J.Paliwal, N.D.G.White; Detection of insect-damaged wheat kernels using near-infrared hyperspectral imaging[J], Journal of Stored Products Research, **45**, 151-158 **(2009)**.

**[14]** Hu Jie; Survey on feature dimension reduction for high-dimensional data[J], Application Research of Computers, **25(9)**, 2601-2606 **(2008)**.

**[15]** Hu Yu; Applied the PCA dimensiona reduction and Quadratic Discriminant Analysis (QDA) on Gene Data Expression Analysis[J], Journal of Guangdong Polytechnic Normal University, **10**, 25-27 **(2007)**.

**[16]** J.H.Friedman; Reguralized discriminant analysis, Journal of the American Statistica Association, **84(405)**, 165-175 **(1989)**.

**[17]** Zhang Feng, Li Hui, Zhang Qin-Yu; A Method for Reducing the Effect of False Alarm and Missing Alarm on HARQ-failure[J], Telecommunication Engineering, **49(8)**, 81-85 **(2009)**.