

# RESEARCH OF RICE-QUALITY BASED ON COMPUTER VISION AND NEAR INFRARED SPECTROSCOPY

RuoKui Chang<sup>1</sup>, WeiYu Zhang<sup>1</sup>, Jing Cui<sup>2</sup>, YuanHong Wang<sup>3\*</sup>, Yong Wei<sup>1</sup>, Yuan liu<sup>1</sup>

<sup>1</sup>*Department of Electromechanical Engineering, Tianjin Agricultural University, Tianjin, P.R. China 300384*

<sup>2</sup>*Department of Agronomy, Tianjin Agricultural University, Tianjin, P.R. China 300384*

<sup>3</sup>*Department of Horticulture, Tianjin Agricultural University, Tianjin, P.R. China 300384*

\*Corresponding author, Address: Department of Horticulture, Tianjin Agricultural University, Tianjin, P.R. China 300384, Tel:13652126792, Fax:86-22-23781291, Email:changrk@163.com

**Abstract:** A rapid and nondestructive way to measure protein and amylose content of rice was put forward based on near infrared(NIR) spectral technology. The NIR spectra were acquired from 13 varieties of rice with the wavelength from 700 to 1100nm. The objectives of the present study were to establish forecasting model to find out the relationship between the absorbance of the spectrum and the main components of rice. By using the machine vision-based method, the rice appearance quality can be studied. On the basis of the evaluation criteria, 13 different kinds of rice were classified. And according to the usage of neural network, the detection model was established, so it can lay the foundation for the prediction grade of the unknown kinds of rice in the future.

**Key words:** near-infrared spectroscopy; appearance quality; machine vision; artificial neural network

## 1. INTRODUCTION

Rice is the major food crops of China. Rice-quality is the basic characteristic of rice as a commodity in the circulating process. It mainly includes internal quality and appearance quality. The amylose content (AC) and the protein content (PC) are the two central indicators of the internal

quality of rice. The amylose content is one of the most important elements affecting the rice cooking and processing properties. The amount of it will directly affect the cooking and eating quality of rice. So it is often used as the evaluation index of grain quality characteristics of the parboiled rice. Protein is the major nutrient of rice, but the rice with high protein content will become light yellow, and can be degenerative during the process of storage. It makes the quality of cooking and appearance of rice lower (Mouhuidong et al., 1993; Liu Jianxue et al., 2001). So the protein content of rice is often regarded as one of the most indexes for the evaluation of grain quality. As a main commodity trait of rice, the appearance quality determines the market price of rice to a large extent. At present, the evaluation method of the rice quality in China is still at the level of naked eye observation. The test results not only lack objectivity and repeatability, but also time consuming and laborious for operation. It has the lower cognition rate of the trait parameters which reflect the appearance quality of rice, such as head milled rice rate, chalkiness degrees. It greatly reduces the competitiveness of China's rice in the international market.

For these reasons, near infrared spectroscopy is used for the building of the forecasting model of amylose and protein content of rice, the non-destructive test of the main components of rice is realized, the testing time is shortened, the internal quality of rice is quickly and easily determined, it is convenient for high-quality breeding; the machine vision technology can be used to identify the form of rice quickly and effectively, and on the basis of the evaluation criteria, different kinds of rice are classified.

## 2. MATERIALS AND METHODS

### 2.1 Preparing test material

The test samples of rice are provided by China-Japan Joint Center on palatability and quality of rice. 13 kinds of samples are selected for the inspection. 20g of each kind of rice is selected; the rice samples can be gained through browning and milling for 2 minutes. Chemical calibration value of amylose and protein content are gained by iodine colorimetry and Kjeldahl method respectively, these values are provided by China-Japan joint center. It is shown in *Fig.1*.

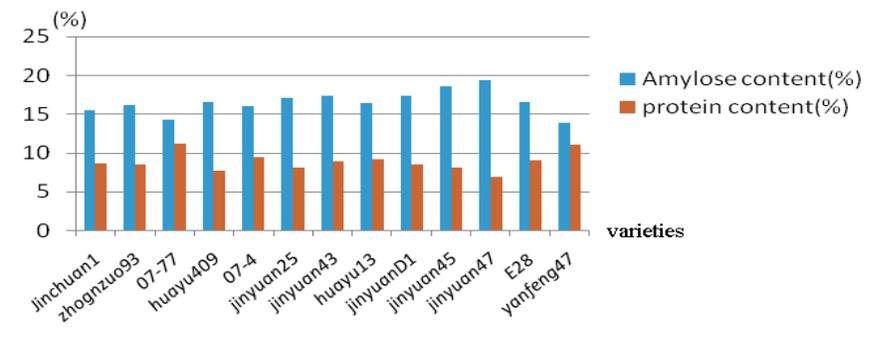


Fig. 1: Amylose content and protein content of different varieties of rice

## 2.2 Non-destructive test of amylose and protein content of rice

### 2.2.1 The composition of NIRS and the testing process

The detection system of amylose and the protein content is made of two parts, computer and near infrared spectroscopy instrument. The USB 2000 miniature fiber optic Spectrometers, which is produced by ocean optics, inc. is used. The wavelength coverage is 200-1100nm. It is mainly composed of the S2000 miniature fiber optic spectrometer, an A/D converter, our operating software, a light or excitation source, and sampling optics.

The testing process is as follows: The light or excitation source sends light through an optical fiber to the single grain. The light interacts with the sample. Then the light is collected and transmitted through another optical fiber to the spectrometer. The spectrometer measures the amount of light and the A/D converter transforms the analog data collected by the spectrometer into digital information that is passed to the software, providing us with application-specific information. Fig.2 shows us the system connection diagram.

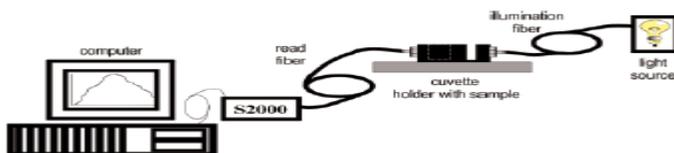


Fig.2: Connection diagram of NIRS detection system

### 2.2.2 Spectral acquisition

The near-infrared spectroscopy of substance is a reflection of its molecular structure. Fig.3 shows the molecular absorption peaks correspond

to the form in the vibration of the rotation group. The absorption peak with XHn (O-H, N-H, C-H) functional groups are in the region domination of the near-infrared spectroscopy (Zhang jun et al., 2003) So three points are selected, which are the frequency characteristics of protein contains, respectively, 785nm (third harmonic generation), 910nm (C-H stretching vibration), 1020nm (N-H stretching vibration, the second octave). The characteristic frequency of amylase is at 990nm (O-H stretching vibration, the second octave) (Li minzan et al., 2006) The peak absorption of Near Infrared Reflectance Spectroscopy of protein and amylase content are collected one after another from 13 kinds of single-grain samples.

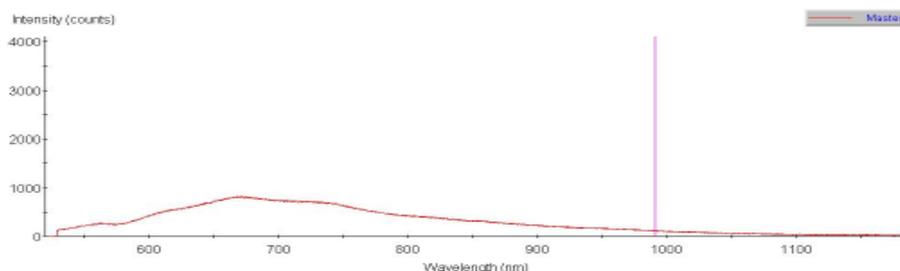


Fig.3: The relationship between absorption peak and wavelength

### 2.2.3 Preprocessing spectral data

Spectroscopy will inevitably produce errors in the determination procedure; the mathematical pre-processing method can be used to deal with the specific samples. The abnormal samples can be removed, spectral noise can be eliminated, data variables can be selected, spectrum of information is purified, and all non-target effects of factors on the spectrum are reduced. It will lay the foundation for establishing spectral correction model and forecasting the concentration of the components of unknown samples.

First of all, a number of points around the smoothing point are selected to get the best estimation of result. Its purpose is to eliminate the random error of high-frequency. Secondly, as a result of the influence of derivative spectra on the elimination of the baseline drifting or the moderating of the background interference, it can also offer the transformation of spectral outline which is higher and clearer than the original spectrum (Zhang jun et al., 2003). Taking the specific circumstances of the laboratory into consideration, we decide to adopt the method of first-order differential to preprocess spectral data. Because the actual measurement of spectrum is discrete spectrum, after the differentiation of it, *table 1* is gained, and it shows the relationship between the absorbance and chemical values at different wavelengths.

Table 1. The relationship between the absorbance and chemical values at different wavelengths

various	Protein					Amylase		
	Content( y)	lgy	lg $\chi_{785}$	lg $\chi_{910}$	lg $\chi_{1020}$	content(y1)	Lgy1	Lg $\chi_{990}$
jinyuan25	8.1	0.908	1.273	0.520	0.318	17.14	1.234	0.420
jinyuanD1	8.6	0.934	0.981	0.417	0.246	17.34	1.239	0.311
07--4	9.5	0.977	1.192	0.502	0.297	16.08	1.206	0.365
jinyuan43	9	0.954	1.216	0.509	0.307	17.42	1.241	0.381
huayu13	9.2	0.963	6.863	0.473	0.463	16.46	1.216	2.363
jinyuan45	8.2	0.913	6.513	0.381	0.080	18.53	1.267	3.813
E28	9.1	0.959	6.584	0.447	0.982	16.59	1.219	2.784
jinyuan47	6.9	0.838	1.060	0.442	0.260	19.42	1.288	0.338
jinchuan1	8.7	0.939	4.392	0.468	2.192	15.45	1.188	2.092
huayu409	7.8	0.892	9.002	0.402	0.430	16.55	1.218	1.969
07-77	11.2	1.049	3.245	0.530	0.645	14.25	1.153	2.012
zhongzuo93	8.5	0.929	1.179	0.475	0.284	16.22	1.210	0.385
yanfeng47	11.1	1.045	8.322	0.404	0.610	13.91	1.143	4.223

#### 2.2.4 The establishment of forecasting models

Near infrared spectroscopy is composed of the absorbance of many wavelengths, the absorbance of each wavelength is the volume, which changes with different specimen. Therefore the spectrum can be regarded as the composition of many variables of multi-variable data. Protein and amylose content of rice are selected as objective functions respectively in this experiment, the absorbance value of absorption data of absorbance band is chosen as independent variables, the partial least squares (PLS) linear regression model of rice is established.

Assuming the prediction model between the rice protein content (y) and absorbance (x) is as follows:

$$y = -0.0144x + 0.9126 \quad (1)$$

Assuming the prediction model between the amylose content (y) and absorbance (x) is as follows:

$$y = -0.0027x + 1.3251 \tag{2}$$

Table 2 shows the results of calibration test of quantitative analysis model of protein and amylase content of rice.

Table 2. The results of calibration test of quantitative analysis model of protein and amylase content of rice

content	Calibration set				Test set			
	Ingredients range	R	SEC	Relative Standard error(%)	Ingredients range	R	SEP	Relative Standard error (%)
amylose	16.08-19.42	0.69	0.03	2.10	13.91-16.55	0.69	0.02	1.93
protein	6.9-9.5	0.61	0.16	7.12	7.8-11.2	0.64	0.04	4.43

Note: SEC is the standard error of estimate from the calibration; SEP is the standard error of prediction.

## 2.3 The establishment of machine vision systems

### 2.3.1 Image sampling of rice

By using color image analysis system, the image sampling of rice of single grain is completed. The color image analysis system is composed of computer system, camera system and color image analysis software package. The objects can be analyzed quantitatively by using computer multimedia. 100 images of single-grain of 13 varieties of rice are selected respectively in this test. Using electron microscopy analyzer, we can detect morphology parameters of rice and analyze the morphological characteristics of rice. These features include: the largest and smallest diameter, ratio of grain length to width, the outline of the complexity, the average gray level, the average optical density, the average transmission rate and so on. After image processing, parameters of these patterns are quantitative.

### 2.3.2 Appearance indicators analysis and data processing

The appearance quality refers to the indicators which are related to the characteristics of rice, such as the chalkiness, translucency, and grain shape. Chalkiness refers to the opaque parts of grains (Li tianzhen et.al.2005), it is a kind of optical feature caused by loose filling and inflating of starch and albumen grains in the endosperm of rice. The chalkiness of rice refers to the percentage of the shares of chalky areas in the projected areas of the whole rice when the chalky rice is placed flatly. It is a defect of the grain structure.

It can influence the appearance of the rice, and also, it can make the grains broken in milling process. Thus, it is an undesirable characteristic of rice, and it has become one of the objects which scientists need to research emphatically and overcome. Chalkiness degree is used to describe the degree of chalkiness. It refers to the parts of chalky areas and the percentage of total areas(Sun Ming et.al.2002). The formula is as follows:

$$\text{Chalkiness ratio (\%)} = (\text{part chalkiness area} / \text{total area of grain}) \times 100\% \quad (3)$$

$$\text{Chalky rice ratio (\%)} = (\text{containing rice grains} / \text{total grains}) \times 100\% \quad (4)$$

The translucency of rice is an indicator to describe the transmission characteristics of grains. Variety is the main factor affecting the translucency of rice. Cultivation and processing conditions also have an impact on it. The grain shape is described with the length, width and the ratio of it, as shown in formula 3. In the past the grain shape was judged as quality indicators, it is more reasonable to judge it as classified parameters currently.

$$\text{length/ width ratio of grain} = \frac{\text{average length of grain (mm)}}{\text{average width of grain (mm)}} \quad (5)$$

In accordance with the national rice quality standard (GB / T 17891-1999), the grade of external characteristics can be divided into three levels, according to head milled rice rate, chalky rice percentage, chalkiness degree, translucency, grain shape etc.

If two or more indicators are not up to standard, but can up to the next rank, they will be reduced to a lower rank; if any of the indicators can't meet the three requirements, it can't be used as high-quality rice. The grade judgment of appearance of rice is shown in [table 3](#). By using electron microscopy analyzer, we can collect and analyze the data of single grain to obtain outside characteristic parameters, as is shown in [table 4](#).

Tabl .3. Indicators of appearance quality

Item	Translucency and gloss	Chalkiness size (%)	Chalky rice ratio (%)	Grain shape
First Class	Translucent glossy >0.7	<5	<5	>1.7
Second-class	Translucent 0.61-0.7	<10	<10	>1.6
Third class	Poor transparency 0.46-0.6	<20	<20	>1.5

Table 4. Analysis of different data characteristics of rice

various	grain length to widthratio(LWR)	Average translucency	Chalky rice percentage(%)	Appearance degree
Jinchuan1	1.530121	0.801545	3	II
zhongzuo93	1.520617	0.710783	8	III
07-77	1.706824	0.785607	21	O
huayu409	1.479549	0.777721	6	III
07-4	2.07547	0.709472	3	I
jinyuan25	1.638583	0.761274	6	III
jinyuan43	2.011645	0.780593	3	I
huayu13	1.639941	0.792892	12	III
jinyuanD1	1.603584	0.714055	9	III
jinyuan45	1.577334	0.772294	14	III
E28	1.929455	0.704889	1	I
yanfeng47	1.665421	0.753477	16	III
jinyuan47	1.738327	0.787858	15	III

Note: O stand for out of high-quality rice.

Table 4 shows the appearance quality of 07-4、jinyuan43、E28 are better than others.

### 2.3.3 The building of forecast models

As a new method of information processing, artificial neural network has highly parallel characteristics and the features of input signals. The mapping relationship between input and output can be learned from it, and associative memory can be integrated. When a particular input signal is given, the appropriate output signal can be renewed. And when a new input signal is given, an appropriate output signal can be generated. As it is difficult to establish a linear relationship between the appearance characteristics of rice and grade, in order to rapidly and accurately determine the grade of rice, we use three-tier BP neural network. We use the main components, like translucency, chalkiness size, grain shape, which are extracted from the images of single grain rice as the input layer of BP network. The output layer corresponds to the level of grading standards, so the mapping relationship can be established from P3 to y1.

The 300 single grain of rice were selected randomly, in which 250 samples for training, 50 as testing samples, Fig.4 shows the network structure.

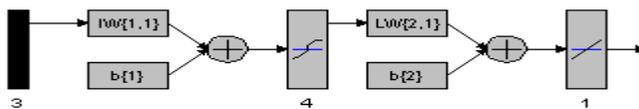


Fig.4: The structure of ANN

The correlation coefficient of the calibration model established by appearance indicators is 0.9621. The finished models are used to forecast the prediction set, the correlation coefficient of the prediction is 0.9134. The calibration set or the prediction set, it can be seen that the correlation between the forecast value and the true value is significant. The using of artificial neural network has a good effect on establishing models.

Table 5 shows the forecast value, the true value and the relative error of 8 samples, which are selected from prediction set.

Table 5. Predicted values and true values of appearance quality degree

subject	Predicted value(degree)	True value (degree)	relative error	subject	Predicted value(degree)	True value (degree)	relative error
1	2.1774	2	0.0887	5	2.1544	2	0.0772
2	2.1544	2	0.0772	6	1.9000	2	-0.05
3	1.1342	1	0.1342	7	1.2198	1	0.2198
4	2.8976	3	0.03413	8	1.7894	2	0.1053

Note: the data listed above are the results of 8 samples among 50 unknown samples

From table 5, we can see that the relative error between the forecast value and the true value is too small. The average relative error is 5%.

### 3 RESULTS AND DISCUSSION

In the experiment, the near-infrared spectral analysis technology is used for the selection of 13 kinds of intact rice samples, the spectrogram of protein and amylase of rice is gathered, By extracting the absorbency from the functional group with the characteristic value of 785nm、910nm、990nm and 1020nm, the forecasting model of quantitative determination of protein and amylose content can be established. In the calibration set, the determination coefficient R of amylose and protein is 0.69 and 0.61 respectively, and the relative standard error is 2.1% and 7.1% respectively. The correlation coefficient r of the validation set is 0.69 and 0.64 respectively. This NIR model is used to predict the contents of amylase and protein in 200 samples, When using the standard error of predication (SEP) to assess the deviation of analysis results is 0.02 and 0.04 respectively, the relative standard error is 1.95% and 4.43% respectively. These results would provide convenience for the breeders to select varieties of high-quality rice to facilitate breeding.

The machine vision technology was used to obtain three appearance parameters of rice. On the basis of the evaluation criteria, 13 kinds of rice

are divided into three grades. And according to the usage of neural network, the detection model is established, the correlation coefficient of the calibration model established by appearance indicators is 0.9621. By using the model to forecast the prediction set, the correlation coefficient is 0.9134. The relative error between the forecast value and the true value is relatively small. The average relative error is 5%, so it can lay the foundation for the prediction grade of the unknown kinds of rice in the future.

## **ACKNOWLEDGEMENTS**

Funding for this research was provided by Tianjin Agricultural University (P. R. China).

## **REFERENCES**

- Guo Wenchuan, Zhu Xinhua. Research Progress of Machine Vision Technology in Grains Identification and classification. *Cereal & Feed Industry*. 2002,6:50-51.
- Li minzan Spectroscopy Analytical Technique and Application. BeiJing: science press 2006.8.
- Li tianzhen, Zhou boqing. Research on rice- quality inspection basing on computer vision technology. *cereal and food industry*. 2005,Vol.12(4):50-53.
- Liu Jianxue, Wu Shouyi ,Fang Ruming. Rapid Measurement of Rice Protein Content by Near Infrared Spectroscopy, *Transactions of The Chinese Society of Agricultural Machinery*. 2001,32(3):68-70
- Sun Ming, Ling Yun, Wang Yiming. Computer Vision Based Rice Chalkiness Detection Using MATLAB. 2002.18(4):146-149.
- Xie Jinchun. The Study of Nondestructive Detection with NIR Spectroscopy in Single Rice Qualities. Master's thesis, China Agricultural University 2003,6:2-4
- Xu Guoqing The rice quality and evolution indicators(1).*Hunan Agricultural*. 1995.2:15.
- Xu Guoqing The rice quality and evolution indicators(1).*Hunan Agricultural*. 1995.3:16.
- Zhang Jun, Zheng Yongmei ,Wang Fangrong. Discussion on some regular methods for cereal near infrared spectra analysis. *Journal of Changchun Post and Telecommunication Institute*, 2003,21(1):4-8