

# Purity Identification of Maize Seed Based on Color Characteristics

Xiaomei Yan<sup>1</sup>, Jinxing Wang<sup>1</sup>, Shuangxi Liu<sup>1</sup>, Chunqing Zhang<sup>\*</sup>

<sup>1</sup>College of Mechanical & Electronic Engineering, Shandong Agricultural University,  
Taian 271018, China

E-mail: <sup>1</sup>{yanxiaomei521, jinxingw, shuangxiliu168}@163.com; <sup>\*</sup>{cqzhang}@sdau.edu.cn

**Abstract.** In order to identify miscellaneous seed from maize seed accurately and rapidly, maize seed purity identification method based on color extracted from the images of both the maize crown and the maize side was proposed for improving maize seed purity. Firstly, segmentation and single extraction were carried on the original image; secondly, the color models RGB and HSV were used to extract multidimensional eigenvectors from the maize crown and the maize side; finally, multidimensional eigenvectors were projected into one-dimensional space through applying Fisher discriminant theory and K-means algorithm was carried on the new color space. The experimental results show that K-means algorithm based on one-dimensional space received through Fisher discriminant theory can effectively identify maize seed purity, and the recognition rate was over 93.75%.

**Keywords:** maize seed, purity identification, color features, Fisher discriminant theory, K-means algorithm

## 1 Introduction

Maize seed purity is one of the fundamental questions affect maize yield. It is estimated that the yield per hectare will decrease about 135 kg if the maize hybrid seed purity drop 1% (Li Liu et al.,2000).At present, there are many methods to identify the maize seed purity, such as the morphology method, seedling morphology, field planting and electrophoresis analysis, etc. These methods require appraiser must have professional quality and the cycle is too long, so they cannot be applied to maize seed circulation field better. The main factors which influence the maize seeds purity are the hybrids from female parent self-cross, impure hybrid seeds and other maize varieties, etc. (in this article, all these impure seeds are all called miscellaneous seeds). Among which the hybrids from female parent self-cross have the largest proportion and their harm is the biggest, so they not only increase the difficulty for field management, prevention and treatment of disease, but also they seriously affect the yield and the quality of maize seeds (ZhongJun Zhang et al., 1997). The impure hybrid seeds and other maize varieties are more easily identified, but the hybrids from female parent self-cross are impossible to be identified by the naked eye due to their color and morphological features are almost the same as normal hybrid seeds.

Image processing technology, to a certain extent, has overcome the above defects; the applications of computer image processing in the fields of maize seed purity test effectively improve the technical level of the seed purity identification. From the 1980s, the researchers at home and abroad used image processing technology to study maize morphology (Tao Song et al.,1996; Liao K et al.,1992; Ni- B et al.,1997; Zhixing Shi et al.,2008), crack(Wenxue Zhu et al.,1998; Junxiong Zhang et al., 2007; Liao K et al.,1993) and color(Liu J et al.,2000; Xiao Chen et al.,2010), and these methods are mostly dominated by morphology, supplied by color, so far, there are no reports about identifying maize purity only with color characteristics. According to the situations, this paper puts forward a method with color space HSV (Hue/Saturation/Value) and RGB (Red/Green/Blue) as multidimensional eigenvectors which were projected into one-dimensional space through Fisher discriminant theory, and K-means algorithm was used to identify maize seed purity.

## 2 Feature Extraction

### 2.1 Machine Vision System

The main body of the paper contains image acquisition system and the image processing system. Image acquisition system is composed of CCD camera, light, shelf and computer, etc. they are shown in figure 1. Image processing system included two parts: image preprocessing, feature analysis and classification.



Fig. 1. Machine vision system

Take normal hybrid seeds 100 and miscellaneous seeds 100 from each species to do the experiment. Figure 2 shows original image of the jundan20 normal hybrid seed.

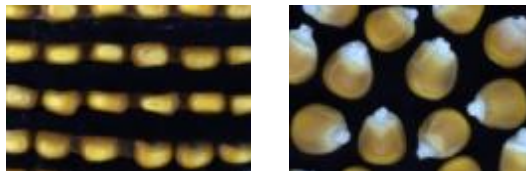
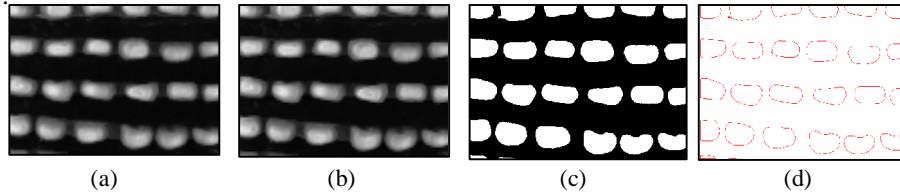


Fig. 2. Original image

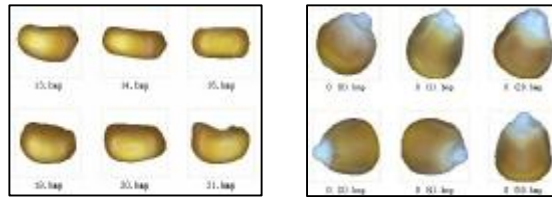
## 2.2 Image Preprocessing

Firstly, 24-bit BMP images were changed into 256 BMP gray images, median filters was used to complete the image enhancement; then the  $R$  plane was used for segmentation. Any pixel with  $R > 55$  was considered as a pixel representing a corn kernel, otherwise, it was considered as a background pixel. Finally, contour labeling was used to extract each seed and display with a color image. Take crown image of Jundan 20 normal hybrid seed as examples, preprocessing results are shown in figure 3.



**Fig. 3.** Preprocessing results image: (a) Gray image. (b) Enhanced image. (c) Background segmentation image. (d) Contour labeled image

The extracted maize crown and no-germ side of the single seed are shown in figure 4.

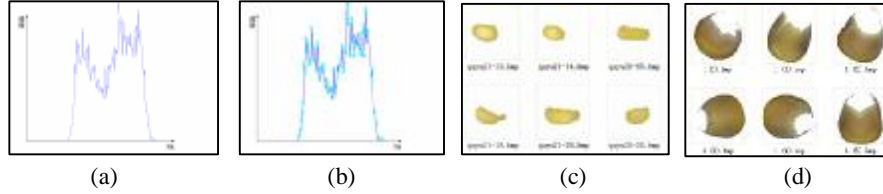


**Fig. 4.** Image of single seed

## 2.3 Region Extraction

Crown center region extraction was completed with regional growing based on gray distribution striving corner detection. Firstly, grayscale histogram and the corresponding gray gradient image were obtained from each seed; secondly, each strong corner on the map gray gradient image was marked and its gray value  $T_i$  ( $i=0, 1, 2, 3, \dots, n$ ) was calculated; then the common values  $T_i$  strong corner which the 60% seeds contains were found out; finally, the strong corner values were sorted from big to small,  $T_0$  and  $T_1$  as optimal threshold for regional growing to segment maize crown into three parts. Namely the region where its threshold was greater than  $T_0$  was the crown center region. The color crown center region can be obtained through comparison of the gray image coordinates and color image coordinates.

The  $B$  plane was used to delete white areas of maize seeds, any pixel with  $B < T$  was considered as a pixel representing yellow region. Otherwise, it was considered as white region. Take jundan20 as examples, gray gradient image, inflexion image, center region image and lateral yellow region are shown in figure5.



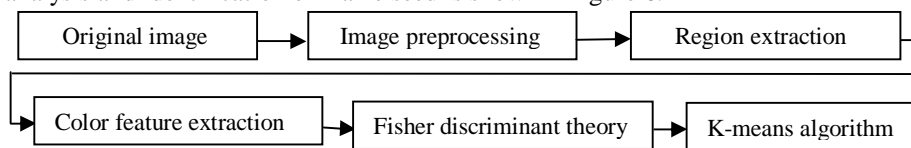
**Fig. 5.** Corner detection and region extraction image: (a) Gradient image. (b) Corner detection image. (c) Crown center region image. (d) Lateral yellow region image.

### 2.4 Color Feature Extraction

In this paper,  $H, S, V, B, G, R$  plane as the color feature were extracted. RGB space is easy for hardware and HSV space is the most suitable for people's vision mechanism. Each plane represented mean of the color features. Due to the range of the  $S$  is  $0 \sim 1$ , which is not able to reflect the change rule, it will be enlarged 100 times in the figure.

## 3 Characteristics Analysis and Identification

The space HSV is evolved by the space RGB, which bring about overlap between  $H, S, V$  and  $B, G, R$ , and each component make contributions to final result of the classification, so if simply to use these color components to classify, it will be difficult to achieve good results. But if the multidimensional color components can be unified one-dimensional space to form a straight line and this line which can integrate each contribution of the components makes each component effect to the maximum; the recognition rate will certainly increase. The flowchart of the characteristics analysis and identification of maize seed is shown in figure 6.



**Fig. 6.** Flow chart of maize seed feature analysis and identification

### 3.1 One-dimensional Function Establishment

According to the above analysis, discriminant function based on Fisher discrimination theory is the line which is required. The basic thought of Fisher discriminant analysis theory is projection. High-dimensional data points are projection to the low-dimensional space, so the data points become more intensive, which can overcome the "dimension curse" arising from the high dimension(Richard A.,et al.,2002).

Discriminant analysis theory can not only classify something, but also can analyze the force magnitude of specific results affected by various factors.

The purity identification in this paper is two types of questions. M dimension characteristic vector was extraction from each class. The general expression straight line is given below

$$y = \sum_{q=1}^m c_q x_q + c_0, \quad (q=1, 2 \dots m; \quad m=1, 2 \dots 6) \quad (1)$$

Where  $x_q$  is color vectors,  $c_q$  is coefficient and  $c_0$  is constant. The function

$$y = \sum_{q=1}^m c_q x_q \text{ is the discriminant which is obtained through Fisher discriminant theory.}$$

In order to get center of mass  $\bar{y}_1$  and  $\bar{y}_2$ , all the sample observations are taken to discriminant formula.

$$\bar{y}_1 = \sum_{q=1}^m c_q \bar{X}_{1q}, \quad \bar{y}_2 = \sum_{q=1}^m c_q \bar{X}_{2q} \quad (2)$$

Where  $\bar{X}_{1q}, \bar{X}_{2q}$  are the mean of the classification,  $\bar{X}_{kq} = \frac{1}{n_k} \sum_{i=1}^{n_k} x_i$ ,

( $k=1,2$ ),  $n_k$  is a one-dimensional feature vector contained the number of data. In order to separate two samples as soon as possible, characteristic analysis function  $y$  is established, and the key of the function is how to find a correct projection direction in which the maize seeds are separated best to get a group of appropriate coefficient  $c_1, c_2, \dots, c_q$  that could project multidimensional space vector  $H, S, V, B, G, R$  into a one-dimensional space, and in this direction the maize seeds are separated best. This is the key to solve characteristic analysis function. In order to obtain this correct projection direction, it must make sure that  $D = \bar{y}_1 - \bar{y}_2$  which is the difference between classes is the maximum and  $V_k = \frac{1}{n_k} \sum_{i=1}^{n_k} (y_{ki} - \bar{y}_k)^2$  which is the difference in class is the minimum.

According to the Fisher criteria, the constructor  $F = \frac{D}{V_1 + V_2}$  must reach maximum.

The points that make  $F$  obtain maximum are the solutions of the equations set which its first-order partial derivatives equal to zero based on the multivariate function for extremum principle and method. Therefore, the following equations can be exported:



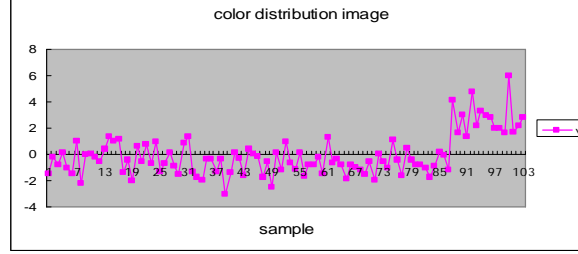


Fig. 8. Color distribution image after Fisher discriminant projection

### 3.2 K-means Clustering

Eigenvector represented by characteristic analysis functions was as the last classification feature. Due to multidimensional eigenvectors were projected into one-dimensional space through Fisher discriminant theory, then k-means algorithm which carried on clustering on the one-dimensional space is simple, intuitive and easy to implement; in addition, the k-means clustering algorithm has the following advantages: k-means algorithm which is a kind of unsupervised statistical method does not need training sample, it makes sample clustering by executing iterative algorithm, its purpose is to make the objective function obtain the minimum. The key to use this method lies in: the data which was obtained through Fisher discriminant has obvious character of classification. The following is the objective function:

$$J = \sum_{j=1}^k \sum_{i=1}^n \left\| x - c_j^{(t)} \right\|^2 \quad (6)$$

Where  $\left\| x - c_j^{(t)} \right\|$  is the distance between data points  $x$  and point cluster center  $c_j^{(t)}$ . Specific steps are as follows:

- Take  $K = 2$  initial center of mass  $C_1, C_2$ .
- All the samples which were obtained in the iteration  $t \geq I$  with the similar assigned to the cluster. Methods: if  $\left\| x - c_1^{(t)} \right\| \geq \left\| x - c_2^{(t)} \right\|$ ,  $x \in G_2^{(t)}$ , where  $G_2^{(t)}$  are the samples by  $c_2^{(t)}$  for clustering center; otherwise,  $x \in G_1^{(t)}$ , where  $G_1^{(t)}$  are the samples by  $c_1^{(t)}$  for clustering center.
- Update clustering center  $c_1^{(t+1)}, c_2^{(t+1)}$  of the clustering  $G_1^{(t)}, G_2^{(t)}$  obtained in step 2, which make  $J = \sum_{j=1}^k \sum_{i=1}^n \left\| x - c_j^{(t+1)} \right\|^2$  ( $j=1, 2$ ) get minimum value.
- For all the  $j = 1, 2$ , if  $x = y$ , iterative end; otherwise,  $t = t + 1$ , turn to step 2.

## 4 Results and Discussion

Some maize seeds (e.g. jundan20, nongda108, zhengdan958, ludan981) were used to verify the experimental method which was proposed in this paper, the results were shown in Table.1. It can be seen from Table1 that the recognition rate is over 93.75% and different varieties have different rate. The reason for the different rate lies in: firstly, the individual differences of corn seed itself is larger: such as the maize crown is full or depression and uneven distribution of color, etc; secondly, the height of the maize crown caused by artificial placing maize seeds is not uniform that causes the differences of the information extracted from the image; finally, clustering features selected in experiments is not perfect, they still need further research. All of these influence region extraction and characteristics analysis in different degree.

**Table 1.** K-means algorithm analysis

Samples	Actual seeds	Miscellaneous seeds	Recognition seeds	Recognition rate/%
Ludan981	100	36	34	94.4
Nongda108	100	32	32	100
Zhengdan958	100	35	34	97.14
Jundan20	100	32	30	93.75

In this paper, a quick and simple method for identify maize seed purity is proposed and the Conclusions obtained through experiments are as follows:

- The purity recognition rate of this method proposed in this paper based on multidimensional eigenvectors  $H$ ,  $S$ ,  $V$ ,  $R$ ,  $G$ ,  $B$  which are projected into one-dimensional space through Fisher discriminant theory and then K-means algorithm is used to cluster analysis is over 93.75% for nongda108, ludan981, zhengdan958, jundan20.
- The crown center region extraction is completed with regional growing based on gray distribution striving corner detection, this method uses the first two strong corner values which the 60% seeds contains as optimal threshold of regional growing and the B plane is used to extract the yellow region. Results show that the areas extracted through the two methods can meet the identification requirements.
- The key of this method lies in how to choose the correct projection direction and the right maize crown center region. In order to verify the validity of this method, a large number of experiments are carried on the other varieties: such as liaoyu19, jinhai601, jinhai702, lainong3, etc., the experimental results are satisfactory.

**Acknowledgments.** This work is supported by Shandong Postdoctoral Innovation Fund(No.72008).



## References

1. Liu, L. and Wang, Y.: Identification of Maize Seed Purity Based on Electrophoresis. Seed World, page 21 (2000)
2. Zhang Z. J., You H. S. and Liu Z. B.: Inbred Seed Harm and Prevention in the Corn Production Field. Jilin Agriculture, page 13 (1997)
3. Cheng H., Shi Z.X., Yao W., Wang L.: Corn Breed Recognition Based on Support Vector Machine. Transactions of the Chinese Society for Agricultural Machinery, 40(3):180-183 (2009)
4. Liao K., Reid J. F. and Ni- B: Corn Kernel Shape Identification by Machine Vision Using a Neural Network Classifier. ASAE Paper, page 7017 (1992)
5. Paulsen M. R., Ni- B and Reid J. F.: Corn Kernel Crown Shape Identification Using Image Processing. Trans of the ASAE, 40(3):833—838 (1997).
6. Shi Z. X., H. Cheng and Li J. T.: Characteristic Parameters to Identify Varieties of Corn Seeds by Image Processing. Transactions of the CSAE, 24(6) : 193—195 (2008)
7. Zhu W. X. and Cao C. W.: Microscopic Structure Analysis on Stress Cracks in Corn. Transactions of the CSAE, 14(2):198-201 (1998)
8. Zhang J. X., Xun Y. and Li W.: Detection of Surface Cracks of Corn Kernel Based on Morphology. Optics and Precision Engineering, 15(6): 951-956 (2007)
9. Liao K. and Paulsen M. R.: Corn Kernel Breakage Classification by Machine Vision Using a Neural Network Classifier. Trans of the ASAE, 36(6):1949-1953 (1993)
10. Liu J. and Paulsen M. R.: Corn Whiteness Measurement and Classification Using Machine Vision. Transactions of the ASAE, 43(3):757-763(2000)
11. Chen X., Xun Y. and Li W.: Combining Discriminant Analysis and Neural Networks for Corn Variety Identification. Computers and Electronics in Agriculture, 71S S48–S53(2010)
12. Richard A. and Johnson D .W. W.: Applied Multivariate Statistical Analysis, Prentice Hall, London (2002)