Quality Classification of Peanut using Computer Vision Technique

Govind Chaudhari[1], Kinjal Mehta[2]

[1,2]Department of Electronics & Communication Engineering,
L D College of Engineering Ahmedabad, Gujarat, India
[1]govind_2121@yahoo.com

Abstract-- **This paper presents a methodology for identification and classification of peanut (Arachis hypogaea L.). Texture, Color and shape features are the basis used for recognition. The texture features are extracted using gray level co-occurrence matrix method and variogram matrix features. Color features are calculated like skewness, kurtosis, mean, and variance. Shape features are calculated using Fourier Descriptor. Neural network is best used for decision making of complex problems. Average accuracy for correlation and ASM parameters is 90%, for correlation and variance is 75%, for variance and ASM is 85%. Approximate accuracy for good peanut is 90% and 85% for bad peanut 80% for moldy peanut using seven texture and one color parameter.**

Keywords -– **Co-occurrence matrix, Skewness, Variance, Kurtosis, Neural network, Variogram**

I. INTRODUCTION

The peanut is high in nutritional value and is an important resource for edible oil and high-quality vegetable protein. It is specifically stipulated in the import and export trade for the peanut appearance quality like damage, mildew, size, shape and colour, etc. Therefore, the classification is an important method to make sure that the peanut can be priced according to its quality. Good-quality peanut deserves a high price. Currently, the photoelectric colour sorter used in the market can only recognize the peanuts of different colour and the impurity, but has no way to detect such quality indices as size, shape, freshness before making a comprehensive judgment, thus difficult to satisfy varied market requirements.

The seed purity of variety and quality in peanut is significant in new variety testing and protecting, seed production and quality control, marketing, imports and exports[4]. The detection of purity of variety in peanut seed mainly concerns the sort of the peanut variety, and the detection of peanut quality mainly scales trade quality especially the sense organs quality. At present, the detection and research of peanut seed carries through mainly with the method of manual work and bio-chemical. The manual method also calls for experience-rich people and the method is fatigable. The bio-chemical identification calls for expensive detecting equipment, precise and complex experimenting technique, and the cost of detecting is high. In order to cope with the demand in peanut breeding, further processing and foreign trade, it is a burning question to find a speedy and accurate method to identify the peanut variety and quality.

The testing method based on image processing and machine vision is a new one which is undamaged, speedy with high distinguishing rate, repeatability and low cost and fatigue; can be used in batch test. The author has applied this method in detecting the variety and quality of a small number of peanut.

We are going to collect a large number of pictures of peanut, design experiment on recognition of peanut variety and peanut quality testing respectively. Then we discuss the feasibility of recognition to peanut appearance characteristic. At last we will analyze the role of appearance characteristic in machine recognition and the factors which affects the result of the detecting.

II. TEST MATERIALS AND IMAGE ACQUISITION

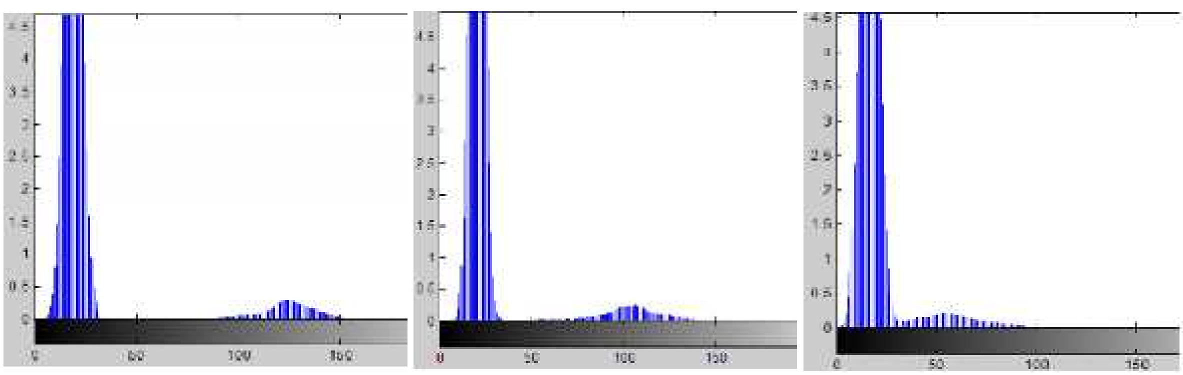
It is found through observation and analysis that peanut with the same variety and batch are mainly classified into five shapes and quality as in Fig 1.Peanuts of above-mentioned are taken 60 kernels each type and then photograph them with SONY digital camera of 10M pixels with the focal length 30mm. All photos are in jpg form.

(a)damage (b)moldy (c)circular (d)elliptic (e)stale

Fig. 1 Types of Selected Peanut

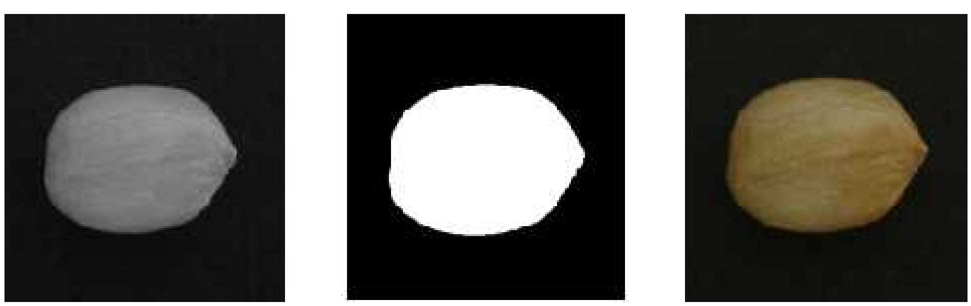
III. IMAGE SEGMENTATION

The R, G, B component histograms were extracted as shown in Fig. 2.For the histogram of red component R (Fig. 2a), the contrast gradient between target and background is the biggest, which is good for the following image processing. Therefore, the red component was used by program to obtain gray image (Fig. 3a). By making adaptive threshold segmentation to the gray image of filtered red component R, the regional image of peanut can be obtained (Fig. 3b). Then it was taken as the template to make AND-operation with R, G, B components separately. Finally the colorized peanut image segmented from the background can be obtained through compositional operations, and the color, texture and contour feature of image are preferably kept as shown in Fig. 3c.



(a) (b) (c)

Figure 2 (a) (b)(c) Component histograms of R, G, B (a R, b G, c B).



(a) R gray scale image (b) Segmented image (c) After composing

Figure 3 Peanuts image processing

IV. COLOR FEATURE EXTRACTION

The values of RGB color components are in the range [0, 1] and Hue (H), Saturation (S) and Intensity (I) components are extracted from these RGB components. The equations (1), (2) and (3) are used to evaluate H, S and I components for a given image sample.

$$H = \cos^{-1}\left\{\frac{\frac{1}{2}[(R-G)+(R-B)]}{[(R-G)^2+(R-B)(G-B)]^{\frac{1}{2}}}\right\} \tag{1}$$

$$s = 1 - \frac{3}{(R+G+B)}[min(R,G,B)] \tag{2}$$

$$I = \frac{1}{3}(R+G+B) \tag{3}$$

The color images are recognized by quantifying the distribution of color throughout the image, change in the color with reference to average/ mean and difference between the highest and the lowest color values. 24 features which reflect color are respectively the mean, variance, skewness, kurtosis of the components of the color space of RGB and HSV which are respectively R, G, B, H, S and V. This quantification is obtained by computing mean, variance and range for a given color image. Since these features represent global characteristics for an image, we have adopted mean, variance and range color features in this work. The equations (4), (5) and (6) are used to evaluate mean, variance and range of the image samples.

$$\text{Mean } \mu = \sum x \sum P(x,y) \tag{4}$$

$$\text{Variance} = \sum (x-\mu)2\, P(x,y) \tag{5}$$

$$\text{Range} = \text{Max}(p(x,y)) - \text{Min}(p(x,y)) \tag{6}$$

V. PEANUT SHAPE RECOGNITION

Shape is an important feature for peanut grading. The shape of peanuts varies and can be mainly divided into four types as common, triangular, circular and elliptical. There are many methods to describe the shape of agricultural products. 8 features reflect the shape such as rectangle degree, ovality, concave-convex proportion, circularity, ratio of the length of the major axis and minor axis, compact ratio, ordinate and abscissa of the opposite centroid. The commonly-used statistical and geometric structure methods are simple and practical; however, they are not suitable to describe the agricultural products with irregular shape due to the low accuracy. It is of high accuracy and speed by using Fourier descriptor to describe the irregular shape. This paper studied on the peanut shape classification based on Fourier descriptor.

The boundary tracking algorithm of 8-neigborhood search is adopted to track the boundary contour curve of peanuts. Discrete
Fourier transform to the radius sequences is obtained by tracking the boundary and called Fourier descriptor of boundary as shown in equation below:

$$F(h) = \sum r_g(k) e^{-j2\pi hk/N} \tag{7}$$

Wherein, N stands for the total number of sample points in a cycle. Harmonic order k = 0, 1, …, N-1 and frequency h = 0, 1, …, N-1.

The relationship between frequency hand amplitude |F(h)| describes the distribution of the amplitude of each harmonic with change of the frequency, which indirectly reflects the shape information of primary boundary curve. It can be seen from the figure that the energy of Fourier descriptor is mainly concentrated close to F(0) and F(n). Most parts of values in the middle are very small, and the amplitude is reduced rapidly with the increase of frequency h. It is known through test that when the inverse transform is done to the first 13 components of Fourier descriptor, the shape of obtained boundary image is almost same with the primary one. Therefore, they are considered to be able to stand for the shape of peanut but with no necessary shape information missed. Extract the Fourier descriptors of shape in peanut image, and use the first 13 harmonic components to form a 13 dimension vector $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \ldots \alpha_{12}, \alpha_{13})$ which can be taken as the feature index input value of shape described in this image.

VI. TEXTURE FEATURE VALUE

The surface of normal peanut is smooth while that of mildewed one is complex, namely, the texture feature is obvious. Five co-occurrence-matrix-based texture features as angular second moment, inertia moment, deficit moment, entropy and correlation were chosen. Wherein, angular second moment is the square sum of each element of gray level co-occurrence matrix, which reflects the uniformity degree of image gray distribution as well as the granularity of texture. The bigger its value is, the coarser the texture is. Inertia moment reflects the definition of image and the depth of texture. The bigger its value is, the clearer the image is, and the more obvious the texture is. Deficit moment measures the local variation of image texture. Entropy stands for the information of image and reflects the complex degree of texture. Correlation describes the similarity degree of space gray-level co- occurrence matrix elements in the direction of line and column.

Here the texture features of three types of peanuts with different freshness are extracted. Through analysis within groups, the other four texture features of these three types of peanuts have obvious differences ($P<0.05$) except the inertia moment, where the difference of correlation is extremely obvious ($P<0.01$). It is known through multiple comparison analysis among groups that the difference significance of the five texture feature indexes can be arranged as: correlation > angular second moment > entropy > deficit moment > inertia moment. Here the correlation was taken as the texture feature index of peanut for its better description ability because it can best reflect the surface texture feature of peanuts of different freshness degree.

The texture features such as Contrast, Correlation, Entropy, Energy, Homogeneity, Dissimilarity, Smoothness, Cluster Shade, Cluster Performance, Angular Second Moment, Third Moment, Mean, Variance, Standard Deviation, and Maximum Probability are extracted using gray level co- occurrence matrix. The extracted features are used to train the developed neural network model. The developed neural network model is tested for recognition and classification of different peanut samples.

VII. GRAY LEVEL CO-OCCURRENCE MATRIX (GLCM) METHOD

GLCM is a two dimensional matrix of frequencies at which two pixels, separated by a certain vector, occurs in the image. i.e., the GLCM is a tabulation of how often different combination of pixel brightness values (gray levels) occur in an image. The distribution in the matrix depends on the angular and distance relationship between pixels. Varying the vector used allows the capturing of different texture characteristics. Once the GLCM has been created, various features can be computed from it. After creating GLCM, it is required to normalize the matrix before texture features are calculated. The measures require that each GLCM cell contain a count, but rather a probability.

To accomplish texture analysis task, the first step is to extract texture features that most completely embody information about the spatial distribution of intensity variations in the textured image. Texture features derived from GLCM using the formula.

Contrast returns a measure of intensity contrast between a pixel and its neighborhood. Contrast is 0 for a constant image. Energy means uniformity or angular second moment (ASM) .The more homogeneous of image, larger the value. When energy equals to 1, the image is believed to be a constant image. Entropy is a measure of randomness of intensity image. Image with more number of occurrences of particular color configurations has resulted in higher value of entropy. Local Homogeneity measures the similarity of pixels. Diagonal gray level co-occurrence matrix gives homogeneity of 1.Cluster Shade and cluster prominence are measures of skewness of the matrix. Maximum Probability gives the maximum occurrence of gray levels.

VIII. VARIOGRAM METHOD

The variogram is related to other statistical approaches like the autocorrelation function and the fractal Brownian motion. On the other hand, it is computationally simple and easy to interpret as a graph. One point in which the variogram appears more appropriate is that only weak stationarity is assumed, in other words, the expectation only has to be constant locally.

It appears, however, that most techniques using the variogram do so in the geostatistical manner, i.e., a model is usually applied whose parameters are taken as a way of describing the semi-variogram curve[18]. In Remote Sensing images some texture based variogram might be best modelled using the *spherical* model while others are best represented with an *exponential* or even *sinusoidal* model.

This poses a problem in terms of creating a systematic approach. One solution would be to use the “best” model type, selected as a texture feature. But using a nominal scale feature would cause problems further down the classification process. This would also imply that a battery of models would have to be fitted for all pixels of all texture samples, and the cost in terms of computing would be high. For these reasons and because others have already pursued that line of research, the “traditional” function representation of *sill* and *range* has not been considered here. Another point that has received attention is the alternate use of the *mean square-root pair difference* (SRPD(h)) function proposed by Cressie and Hawkins (1980) as a semi-variance estimator which is resistant to outliers. Lark (1996) has also shown that for four different classes of texture (urban, farmland, woodland, and meadow), when tested for normality, the SRPD(h) function scored much better than the g(h) function, a fact confirmed by an earlier study by the author.

After numerous tests using different ways to transform the variogram into texture features, the most promising approach was found to be the averaging of selected distance lag intervals. The SRPD texture feature extraction routine can be summarized in the following steps:

• For every pixel in the image, a neighbouring window (32 by 32 pixels) is considered and four directional variogram (0°, 45°, 90°, and 135°) are computed for all possible combinations in that window.
• The maximum lag size is equal to one half the window size.
• The *mean Square-Root Pair Difference* is used as semi-variogram estimation.
• Six values (features) of the SRPD are computed by giving more weight to the values corresponding to the smaller lags; in other words, by computing the SRPD features over regular intervals on a logarithmic scale.
• These values are computed for all four directions for a total of 24 features.

The 24 directional features are then transformed to 18 rotation invariant features: for each lag, the *mean*, *standard* *deviation* and *sum of perpendicular ratios* $\sum\left[\frac{\gamma_0}{\gamma_{90}}+\frac{\gamma_{90}}{\gamma_0}+\frac{\gamma_{45}}{\gamma_{135}}+\frac{\gamma_{135}}{\gamma_{45}}\right]$ Where, γ is the estimate of variance) are computed.

Where, the latter two parameters are meant to preserve anisotropy in the data.

IX. NEURAL NETWORK MODEL

The multilayer feed forward neural network model with back propagation algorithm for training is employed for classification task as shown in Figure 4. The number of neurons in the input layer is equal to the number of input

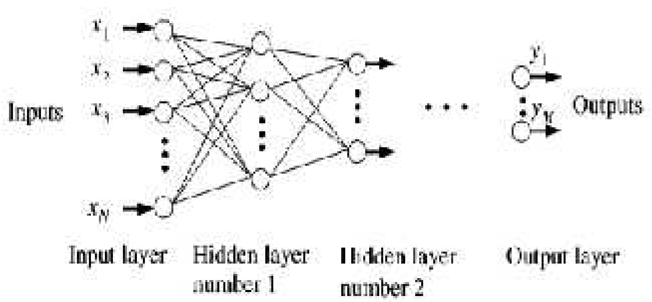


Figure 4 Neural network

features. The number of neurons in the output layer is equal to the number of categories of peanut types. The texture features of an image are used as the inputs.

X. RESULTS AND DISCUSSION

TABLE 1 Result of some quality features of three grades of peanut

	Good	Bad	Moldy
ASM	0.968	.658	.783
Correlation	23127950	33206688	10120050
Contrast	41.35	62.28	54.3
Entropy	21.93	23.27	20.32
Skewness	-.1.5535	-2.1404	-3.543

TABLE 2 Summary of recognition result of peanut qualities

Classification Parameter	Accuracy (%)		
	Good	Bad	Moldy
Correlation	60	80	75
Asm	60	60	80
Variance	30	90	85
Correlation & Asm	70	90	75
Correlation & Variance	60	80	70
Variance & Asm	70	70	75
Corr, Var & Asm	70	70	85

TABLE 3 Result of Peanut Quality

Model	Shape	Texture	Color	Combination
Good	52	48	93	95
Bad	71	84	91	93
Moldy	59	76	94	96

With the detection result we get this points: (1) Different seed appearance characteristic has different role in detection, the detecting result shows shape < texture < colour character; (2) The colour character of the seeds is very obvious in detection, it will improve the detection result by adopting other character parameters; (3) The discrimination validity of appearance characteristic in variety detection is smaller than in quality detection. The method of image processing is more adequate in peanut quality detection.

It should be pointed that the more varieties, the more overlaps in characteristics and discrimination validity become worse, influence the recognition result. The number of the seeds reflexes the representativeness to the variety. Thus combination of shape, texture and colour features gives good result.

Variogram features are included together with GLCM features. Variances parameter of GLCM only includes variance at one pixel distance. That is why GLCM variance parameter is not of much use directly. While variogram calculates variances at every possible lag distance. So adding variogram feature parameters improves result.

XI. CONCLUSION

Selecting features of an image for peanut classification requires thorough knowledge and careful investigation. Texture and color features show different degree of sensitivity towards peanut quality. Combining different features for peanut classification affects overall efficiency. Average accuracy for correlation and ASM parameters is 90%, for correlation and variance is 80%, for variance and ASM is 80%. By combining VGM and GLCM texture parameters in neural network result accuracy is improved. Shape parameters are included to detect bad peanut.

Approximate accuracy for good peanut is 80% and 85% for bad peanut using three texture and one color parameter.

REFERENCES

[1] Chen Hong, Zeng Chuanhua, Ding Youchun,“ Peanut Shape Recognition Based on Fourier Descriptor”, Wuhan, China.

[2] HAN Zhong-zhi; LI Yan-zhao, “Quality Grade-Testing of Peanut Based on Image Processing”, *2010 Third International Conference on Information and Computing, Qingdao, China.*

[3] Basavaraj .S. Anami and Vishwanath.C.Burkpalli, Texture based Identification and Classification of Bulk Sugary Food Objects” *ICGST-GVIP Journal, ISSN: 1687-398X, Volume 9, Issue 4, August 2009*

[4] Hong Chen , Jing Wang, “Quality classification of peanuts based on image processing” Journal of Food, Agriculture & Environment Vol.9 (3&4): 205-209. 2011

[5] Robert Haralick, K Shanmugam, “Texture Feature for image classification” *IEEE transactions on systems,November,1973*

[6] Neelamma K. Patil, Virendra S. Malemath, “Color and Texture Based Identification and Classification of food Grains using different Color Models and Haralick features” *International Journal on Computer Science and Engineering (IJCSE) Dec 2011*

[7] Blasco, J., Cubero, S. and Gómez-Sanchís, J., Mira P. and Moltó E. 2009.‘’Development of a machine for the automatic sorting of pomegranate (*Punica granatum*) arils based on computer vision”. Journal of Food Engineering 90(1):27-34

[8] Neelamma K. Patil and Ravi M. Yadahalli, “Classification of food grains using HSI color model by combining color and texture information without performing pre-processing and segmentation” World Journal of Science andTechnology2012,2(5):50-53

[9] HAN Zhong-zhi and ZHAO You-gang, “A Method of Detecting Peanut Cultivars and Quality Based on the Appearance Characteristic Recognition” 2009 Second International Conference on Information and Computing Science

[10] Tadhg Brosnan and Da-Wen Sun, “Improving quality inspection of food products by computer vision” Journal of Food Engineering 61 (2004) 3–16

[11] J. Sundaram ,C. V. K. Kandala and C. L. Butts, “Classification of in-shell peanut kernels non destructively using VIS/NIR reflectance spectroscopy” Sens. & Instrumen. Food Qual. (2010) 4:82–94

[12] James I. Davidson, Jr.,Paul D. Blankenshipl, and Victor Chew, “Probability Distributions of Peanut Seed Size”, Peanut Science (1978) 5, 91-00

[13] Muthukrishnan.R and M.Radha, “Edge Detection Techniques for Image Segmentation”, International Journal of Computer Science & Information Technology (IJCSIT) Vol 3, No 6, Dec 2011

[14] Andrzej Materka and Michal Strzelecki, “Texture Analysis Methods”, COST B11 report, Brussels 1998

[15] Hong-Lian Li, Xiao-Ting Li, Zhi-Lei Zhao And Yan-Ping Pang, “Cluster Analysis Method and Near-Infrared Spectroscopy Applied to the Identification of Food”, Proceedings of the Ninth International Conference on Machine Learning and Cybernetics,Qingdao (2010),11-14

[16] Jayashree B., Morag Ferguson and Dan Ilut, “Analysis of genomic sequences from peanut”, Electronic Journal of Biotechnology ISSN: 0717-3458 Vol.8 No.3, Issue of December 15, 2005

[17] Comparing Texture Analysis Methods through Classification by Philippe Maillard, 2009

[18] Texture Analysis Methods – A Review by Andrzej Materka and Michal Strzelecki, 2010

.