International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Principal Component Analysis Based Classification Technique for Basmati Rice Grain Analysis

Rubi Kambo[1], Amit Yerpude[2]

[1]M.Tech Scholar, Department of Computer Science and Engineering, CSVTU University
Rungta College of Engineering & Technology, Bhilai (C.G) India

[2]Associate Professor, Department of Computer Science and Engineering
Rungta College of Engineering & Technology, Bhilai (C.G) India

Abstract: ***Different quality and variety of basmati rice available in market, which classification is done on the basis of size and texture features. This paper proposed a new method based on component analysis over extracted shape features of rice granule sample. This method is used for the classification of different variety of Daawat Rozana Basmati rice Gini, Mini and Tini by using a dimensionality reduction and classifier approach. The Experimental result shows the effectiveness of this method for different variety of Daawat basmati rice.***

Keywords**:** Image pre-processing, Principal Component Analysis, Feature extraction

1. Introduction

Rice is the seed of the monocot plants Oryza sativa (Asian rice) [1] [2] [3]. In Asia, especially as a cereal grain, it is the most widely consumed staple food for a large part of the world's human population [4]. After corn, it is the grain with the second-highest worldwide production, according to data for 2010 "Prod STAT" by FAOSTAT (Food and Agriculture Organization of the United Nations) Retrieved December 26, 2006).

Rice is the staple food of over half the world's population. For 17 countries in Asia and the Pacific, 9 countries in North and South America and 8 countries in Africa, it is the predominant dietary energy source. Rice provides 20% of the world's dietary energy supply, while wheat supplies 19% and maize (corn) 5%. ("Rice is Life". Food and Agricultural Organization of the United Nations. 2004.).

Basmati rice means the "queen of fragrance or the perfumed one." According to the Agricultural and Processed Food Products Export Development Authority (APEDA), India is the second largest producer of rice after China, and grows over a tenth of the world's wheat. In 1993, Basmati rice attracted the highest premium because it is very-long grained rice, with an aroma of its own which enhances the flavors it's mixed with.

"Basmati" is long grain aromatic rice grown for many centuries in the specific geographical area, at the Himalayan foot hills of Indian sub-continent, blessed with characteristics extra- long slender grains that elongate at least twice of their original size with a characteristics soft and fluffy texture upon cooking, delicious taste, superior aroma and distinct flavor, Basmati rice is unique among other aromatic long grain rice varieties. Having an average length of about 7.00mm [5] [6]. An excellent quality Indian Basmati Rice that includes Kohinoor Basmati Rice, India Gate Basmati Rice, Daawat Basmati Rice etc.

To satisfy the taste buds of consumers, the organization conducted an extensive consumer research. Upon analysing the findings of the research, four new variants of Daawat Basmati Rice in the market namely Daawat Biryani Basmati Rice, Daawat Traditional Basmati Rice, Daawat Super Basmati Rice and Daawat Pulav Basmati Rice are launched. All these four variants are launched exclusively for Indian market and cater to popular Indian dishes viz. Biryani, pulav, aromatic plain white rice and basmati rice for daily consumption. Daawat variants were launched in India in January 2011 in innovative packaging and they have become people's favourite in such a short span of time.

During the process of rice milling the grain tends to break, these broken grains are separated and packed accordingly on the basis of their size. Although Rozana Gold is a head grain which has all the goodness of pure Basmati but is not long enough to make it to the premium category, all other variants are broken and thus named on the basis of the grain length.

Variants of Daawat are ROZANA BASMATI RICE GINI 70, ROZANA BASMATI RICE MINI 60 and ROZANA BASMATI RICE TINI 50.

In pattern recognition and image databases, the data is made up of a set of objects, and the high dimensionality is a direct result of trying to describe the objects via a collection of features (also known as a feature vector). Examples of features include color, color moments, textures, shape descriptions.
Dimensionality reduction or dimension reduction is the process of reducing the number of random variables under consideration and can be divided into feature selection and feature extraction.

Feature extraction transforms the data in the high-dimensional space to a space of fewer dimensions. The data transformation may be linear, as in principal component analysis (PCA), but many nonlinear dimensionality reduction techniques also exist.

Licensed Under Creative Commons Attribution CC BY

2. Principal Component Analysis

For dimensionality reduction, principal component analysis the main linear technique performs a linear mapping of the data to a lower-dimensional space in such a way that the variance of the data in the low-dimensional representation is maximized. In practice, the correlation matrix of the data is constructed and the eigenvectors on this matrix are computed. The eigenvectors that correspond to the largest eigenvalues (the principal components) can now be used to reconstruct a large fraction of the variance of the original data. Moreover, the first few eigenvectors can often be interpreted in terms of the large-scale physical behaviour of the system. The original space (with dimension of the number of points) has been reduced (with data loss, but optimistically retaining the most important variance) to the space spanned by a few eigenvectors [7].
Principal component analysis (PCA) is a classical statistical method. This linear transform has been widely used in data analysis and compression. Principal component analysis is based on the statistical representation of a random variable [7]. PCA summarize the data as follows:

Step-1-Take an original data set and calculate mean of the data set taking dataset as column vectors, of M rows. Then Place the column vectors into a single matrix X of dimensions M × N.

Step-2- Subtract off the mean for each dimension. Place the calculated mean values into an empirical mean vector u of dimensions M × 1.

$$u[m]=(1/N)\sum_{n=1}^{N} X[m,n]$$

Mean subtraction is way of finding solution toward principal component ,that minimize the mean square error and then approximating toward the data and for this, two steps are:

1. Subtract the empirical mean vector u from each column of the data matrix X.
2. Store mean-subtracted data in the M*N matrix B.

$$B=X-uh$$

[Where h is a 1 x N row vector of all 1's: h[n] =1 for n=1…N]

Step-3- Calculate the covariance matrix. By using the formula,

$$C=E[B\times B]=E[B.B^*]=(1/N)\sum B.B^*$$

Step-4- Calculate Eigen vector and Eigen value of the covariance matrix. Compute the matrix V of eigenvectors which diagonalizes the covariance matrix

$$C:\ V\text{-}1CV=D$$

Where, D is the diagonal matrix of eigenvalues of C.

Step-5- Extracting diagonal of matrix as vector.

$$D[p,q]=\lambda m$$

Step-6-Now, Sort the column of the Eigen vector matrix V and Eigen value matrix D in order of decreasing Eigen value i.e sorting invariance in decreasing order.

Step-7-Choosing components and forming a feature vector that is notion toward data compression and dimensionality reduction.

In fact, it turns out that the eigenvector with the highest eigenvalues is the principle component of the data set. And finally we need to form a feature vector.

Feature Vector = (eig1 eig2 eig3…….eign)

Step-8- Finally, Deriving g the new dataset.
Once the components (eigenvectors) that we wish to keep in our data is chosen and a feature vector is formed, we simply take the transpose of the vector and multiply it on the left of the original data set, transposed.

Final Data=Row Feature Vector*Row Data

where Row Feature Vector is the matrix with the eigenvectors in the columns transposed so that the eigenvectors are now in the rows, with the most significant eigenvectors at the top, and Row Data Adjust is the mean-adjusted data transposed, i.e. the data items are in each column, with each row holding a separate dimension [8].

3. Materials and Methodology

From Retail store, basmati rice grain samples of Daawat brand were collected and Samsung Galaxy mobile camera is used to acquire and record the images .For this rice grain is placed on the black sheet of paper and the camera is placed at fixed location and mounted on stand, to provide vertical movement and the distance between the lens and the sample table is 14 cm .All the grains in the sample image were arranged in arbitrary direction and position .The Black background is used and light intensity on sample table is uniform to improve the data collection under controlled environment. . Images were captured and stored in JPG format automatically .Through data cables these images has been transferred and then stored in hard disk and then different parameters of rice were extracted from the image for further analysis [9] .

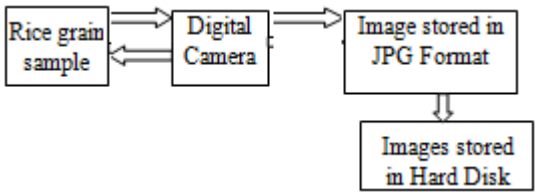

Figure 1: Basic Building block for image capturing [9] [10]

4. Image Analysis

4.1 Image Acquisition

The Image acquisition is first step in image processing because without an image, no processing is possible ,thus the image acquisition is the process of acquiring an image from a source such as camera or scanner .Acquisition of image can be done under uniform lighting by Samsung mobile Digital camera [11] [10] [12].

Licensed Under Creative Commons Attribution CC BY

4.2 Image Enhancement and Smoothing

While capturing the image, sometime it has been distorted and hence image is to be enhanced. Preprocessing aim is to improve an image and suppress the unwanted distortion and enhance image feature for further processing. Noise can be removing from image by applying filters. Noise reduction techniques such as Averaging, Gaussian filters are used and causes image smoothing .In this paper, Median filter is used for smoothing because it protect the edges of the image during noise removal and is mostly used in digital imaging and effective with salt and pepper noise and speckle noise .The noise in the input gray color image is detached using median filter.[13] [11].

4.3 Image Segmentation

The next process in image processing is the image segmentation after the acquisition and the very first step in image analysis is image segmentation where the image is subdivided into different parts or object. Basically the image is subdivided until we segregate the interested object from their background. Generally there are two approaches for segmentation algorithms. one is based on the discontinuity of gray level values and the other is based on the similarity of gray level values and for this different approaches like thresholding, region growing ,region splitting and merging can be used [13] [11] [10] [14].

Segmentation can also be done using edge detection. The simplest method of image segmentation is called the thresholding method. By using threshold value, image binarization is performed. Threshold is used to separate the region in an image with respect to the object, which is to be analyzed and this is based on the variation of intensity between the object pixel and background pixel.

Another approach is a region growing method used for segmentation .In the present research work after enhancement of image, the region of each rice grain in an image is detected using region growing. More specifically, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics [11] [13] [15] [16].

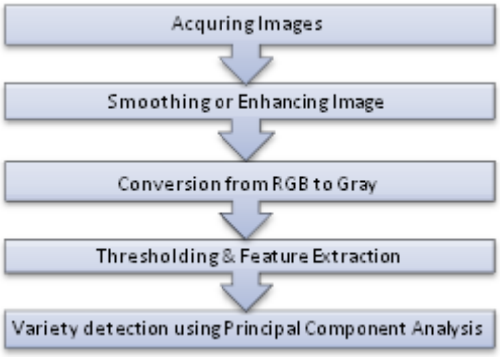

Figure 2: Flow chart for general steps

4.4 Feature Extraction

Feature Extraction deals with Extraction of quantitative information from image. The work of categorization and recognizition (identification) of object based on various feature such as morphological feature, color feature extraction and textural feature [17] [18] [19] [20] [21]. In the present research work morphological feature are extracted.

Algorithms were developed in windows environment using MATLAB 7.7.0 programming language to extract morphological features of individual basmati rice grains. The following morphological features were extracted from images of individual basmati rice grains:

- **Area**: The algorithm calculates the actual number of pixels inside and including the seed boundary (mm2/pixel).
- **Major Axis Length:** It was the distance between the end points of the longest line that could be drawn through the seed. The major axis endpoints were found by computing the pixel distance between every combination of border pixels in the seed boundary.
- **Minor Axis Length:** It was the distance between the endpoints of the longest line that could be drawn through the seed while maintaining perpendicularity with the major axis.
- **Eccentricity:** specifies the eccentricity of the ellipse that has the same second-moments as the region. The eccentricity is the ratio of the distance between the foci of the ellipse and its major axis length. The value is between 0 and 1.
- **Perimeter:** It was the total pixel that constitutes the edge of object. It helps in locate the object & provide information about the shape of the object i.e. counting the number of '1' pixel that have '0' pixel of neighbour.

5. Result and Discussion

5.1 Image Sample

Sample image of Daawat basmati rice and feature extracted from grain is shown in Table 1.

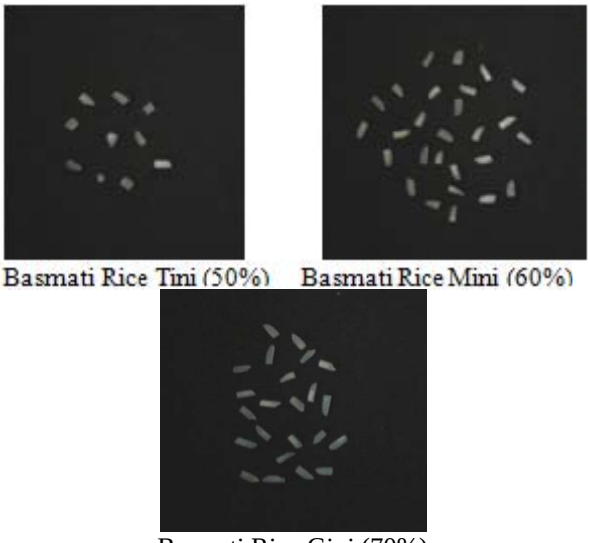

Figure 3: Sample of Daawat Rozana basmati rice variety

Licensed Under Creative Commons Attribution CC BY

Table 1: Features Extracted from Tini

S.no	*Area*	*MajorAxis*	*MinorAxis*	*Eccentricity*	*Perimeter*
1	1795	52.8803	44.404	0.543	161.279
2	1613	67.5933	31.8216	0.8823	191.723
3	1974	70.6091	37.1766	0.8502	185.622
4	836	34.4302	31.0716	0.4308	107.498
5	1912	59.9494	42.7334	0.7013	176.853
6	1978	75.435	34.4978	0.8893	189.179
7	1843	56.0832	42.8227	0.6457	165.966
8	1964	71.3729	36.1134	0.8625	181.622
9	1505	48.4766	41.0192	0.5329	153.137
10	2301	78.3786	38.1192	0.8738	195.498

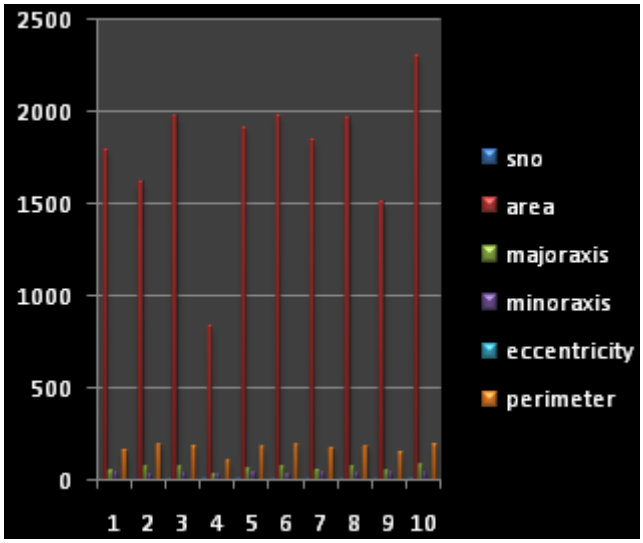

Figure 4: Different Feature Extracted from Sample of Daawat basmati rice variety

5.2 Sample of final data of Daawat Basmati Tini, Mini and Gini rice grain

Table 2: Sample Table for Daawat variety

Basmati Variety	*S.No*	*No. of grains*	*Avg*	*Time*
GINI 70	s1	32	-7.11E-15	3.121105
	s2	44	-2.33E-13	3.139219
	s3	96	1.80E-13	3.174045
MINI 60	s1	46	1.68E-13	3.087053
	s2	69	-1.01E-13	3.069992
	s3	94	-2.29E-13	3.081344
TINI 50.	s1	33	3.79E-14	3.07042
	s2	67	-2.87E-14	3.075596
	s3	113	8.17E-14	3.180843

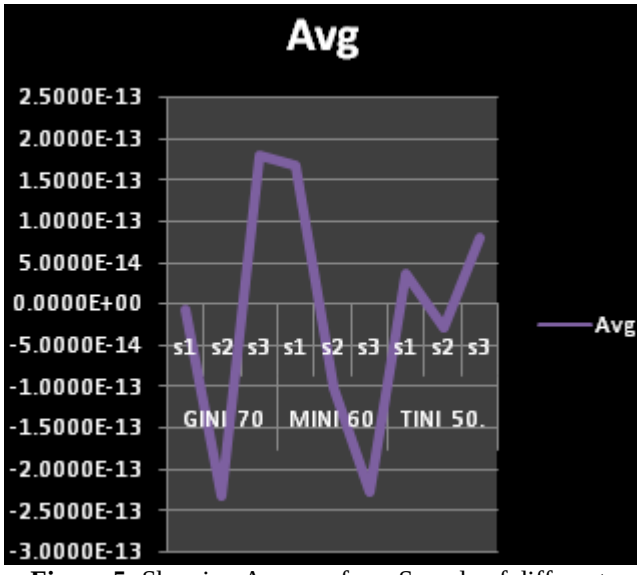

Figure 5: Showing Average from Sample of different Daawat

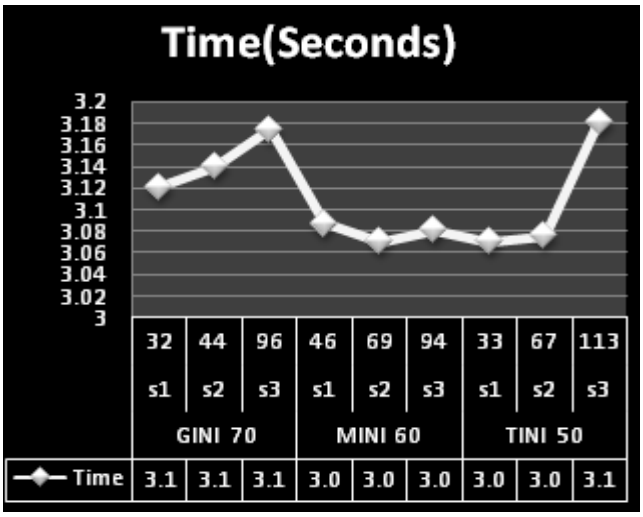

Figure 6: Showing Time Taken by each sample for classification and feature Extraction of rice variety

Table 3: Test Table for Daawat variety

Basmati variety	*Sample no*	*No. of grain*	*Output*		
			Gini	*Mini*	*Tini*
Gini (70%)	GTs1	32	√		
	GTs2	85	√		
	GTs3	97	√		
	GTs4	105		√	
	GTs5	95		√	
Mini (60%)	MTs1	94		√	
	MTs2	95	√		
	MTs3	69		√	
	MTs4	102		√	
	MTs5	30		√	
Tini (50%)	TTs1	67			√
	TTs2	68			√
	TTs3	52		√	
	TTs4	33			√
	TTs5	12			√

Licensed Under Creative Commons Attribution CC BY

Table 4: Accuracy Table for Daawat variety

Variety	*Total no of sample*	*Accuracy*
Gini (70%)	5	60%
Mini (60%)	5	80%
Tini (50%)	5	80%
Overall	15	73%

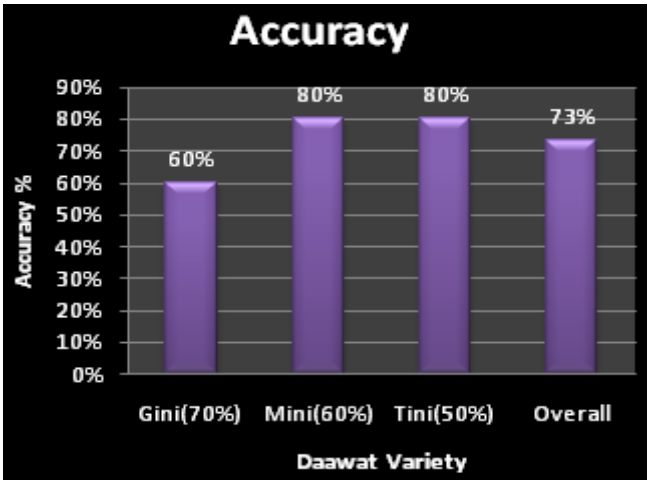

Figure 7: Showing overall and individual Accuracy achieved for classification of Daawat rice variety Sample

According to Table II, it is observed that the value of Principal component is depending upon the number of rice grains present in sample. For classification we use three average values to calculate final value. And for classification K-NN Classifier is used.

Table III and IV shows the output and the effectiveness of proposed method.

6. Conclusions and Future Scope

The Experimental result shows that the proposed algorithm work effectively for samples containing average and more than average number of rice grain for Tini, mini and Gini variety of Daawat basmati rice. The accuracy of proposed algorithm for Tini 80%, Mini 80% and Gini 60%, and the overall accuracy of Daawat basmati rice is 73%. For Future we can improve this system to fully automated system for rice granule grading and detection of adulteration in rice granule. This system can be used for quality assessment of other food granule.

7. Acknowledgement

We would like to express our special thanks to Mr. V C Jain, head of retail shop for the sample of Daawat basmati rice variety as per the requirement desired by us and also thankful for their helpful advice.

References

[1] Amit Yerpude Rubi Kambo, "Classification of Basmati Rice Grain Variety using Image Processing and Principal Component Analysis," IJCTT (International Journal of Computer Trends and Technology), vol. 11, no. 2, pp. 80-85, May 2014.

[2] Chetan V.Maheshwari, "Machine Vision Technology for Oryza Sativa L.(RICE)," IJAREEIE, vol. 2, no. 7, pp. 2893-2900, july 2013.

[3] Chetan V.Maheshwari, "Quality Assesment of ORYZA SATIVA SSP INDICA (Rice)using computer vision," IJIRCCE, vol. 1, no. 4, pp. 1107-1115, june 2013.

[4] D.Swain ,Sanjukta Das,S.G.Sharma & O.N.Singh H.N.Subudhi, "Studies on Grain Yield,Physico-Chemical and Cooking Characters of Elite Rice Varieties(Oryza Sativa.L)in Eastern India," Journal of Agriculture Science, vol. 4, no. 12, pp. 269-275, 2012.

[5] Sachin Patel Bhavesh B.Prajapati, "Classification of Indian Basmati Rice using Digital Image Processing As per Indian Export Rules," International Research Jornal of Computer Science Engineering and Application, vol. 2, no. 1, pp. 234-237, january 2013.

[6] Sachin Patel Bhavesh B.Prajapati, "Algorithmic approach to quality analysis of Indian Basmati Rice using Digital Image Processing," IJETAE (ISSN 2250-2459,ISO 90001:2008certified journal), vol. 3, no. 3, pp. 503-504, March 2013.

[7] Yaduvir singh and sweta tripathi Pramod kumar pandey, "Image Processing using Principal Component Analysis," International Journal Of Computer Application (0975-8887), vol. 15, no. 4, pp. 37-40, february 2011.

[8] Abhishek Banerjee, "Impact of Principal component analysis in the application of image processing," IJARCSSE(ISSN: 2277 128X), vol. 2, no. 1, January 2012.

[9] V.K.Banga Jagdeep singh Aulakh, "Percentage Purity of Rice Samples by Image Processing," in International conference on Trends in Electrical ,Electronics and Power Engineering(ICTEEP 2012), Singapore, july 15-16,2012, pp. 102-104.

[10] Mrs.Uzma Ansari sanjivani shantaiya, "Identification of Food grains and its quality using pattern clasification," vol. 2 , no. 2,3,4, pp. 70-73, december 2010.

[11] Abirami.S,Vishnu Priya.K,Rubalya Valantina.S. Neelamegam.P, "Analysis Of rice granules using Image Processing and Neural Network," in IEEE conference on Information and Communication Technologies(ICT2013), 2013, pp. 879-884.

[12] M.Shahid Naweed ,M NAdim Asif and S.Irfan Hyder Imran S.Bajwa, "Feature based Image Classification by using Principal Component Analysis," CIST-journal of Graphica,vision and image processing.

[13] Kavindra R.jain,Chintan K.Modi Chetan v.Maheshwari, "Non Destructive Quality Analysis of India Gujrat-17 ORYZA SATIVA SSP INDICA(Rice)using Image Processing," IJCES ISSN:22550:3439, vol. 2, no. 3, pp. 48-54, march 2012.

[14] B.Karthikeyan R.Yogamangalam, "Segmentation Technique Comparision in Image Processing," IJET(ISSN :0975-4024), vol. 5, no. 1, pp. 307-313, feb-march 2013.

[15] Gurpreet Singh,Amandeep Kaur Pooja Sharma, "Different Technique Of Edge Detection In Digital Image Processing," IJERA(ISSN:2248-9622), vol. 3, no. 3, pp. 458-461, May-Jun 2013.

[16] G.Kishorebabu,K.sujatha M.Kalpana, "Extraction of Edge Detection using Image Processing Technique," IJCER, vol. 2, no. 5, pp. 1562-1566, september 2012.

Licensed Under Creative Commons Attribution CC BY

[17] Dr.M.Siddappa Harish S Gujjar, "Amethod for identification of rice grainn of india and its quality using pattern classification," IJERA ISSN:2248-9622, vol. 3, no. 1, pp. 268-273, january-february 2013.
[18] Dayanand SAVAKAR, "Recognition and Classification of similar looking food grain images using Artificial Neural Networks," Journal of Applied computer Science & Mathematics , no. 13, pp. 61-65, June 2012.
[19] Chathurika sewwandi silva and upul sonnadara, "Classifiction Of Rice Grain Using Neural Network," in Proceeding of Technical session , SriLanka, 2013, pp. 9-14.
[20] R.M.Carter, Y.yan D.M.Hobson, "Characterisation and Identification ofRice Grains through Digital Image Analysis," in Technology Conference-IMTC, warsaw,Poland, 2007, pp. 1-5.
[21] Virendra S.Malemath,Ravi M.Yadahalli Neelammma K.Patil, "Color And Texture based Identification and Classification of food grains using different Color models and Haralick features," International journal on Computer Science and Engineering(ISSN:0975-3397), vol. 3, no. 12, pp. 3669-3679, 12 december 2011.
[22] Aysegul akdogan Eker Bulent Eker, "Opportunities to use Image Processing Technology in Quality Based Practices," in International Quality Conference, 2013, pp. 67-72. Xx

Author Profile

Rubi Kambo is a M.Tech Scholar of RCET, Bhilai (C.G), India. She did his M.C.A (Master of Computer Application) from Punjab Technical University at Jalandhar (Punjab), India.

Mr. Amit Yerpude is working as Associate Professor in Department of computer Science & Engg. , RCET, Bhilai (Chhattisgarh), India. He has received M.Tech in Computer science from CSVTU University, Bhilai (C.G).He has published much research paper in international journals and presented several research papers in international conferences.

Licensed Under Creative Commons Attribution CC BY