

Automatic Seed Classification by Shape and Color Features using Machine Vision Technology

Naveen Pandey
CSE Department
ASET, Amity University
Noida, India

Satyanarayan Krishna
CSE Department
ASET, Amity University
Noida, India

Shanu Sharma
CSE Department
ASET, Amity University
Noida, India

Abstract: : In this paper the proposed system uses content based image retrieval (CBIR) technique for identification of seed e.g. wheat, rice, gram etc. on the basis of their features. CBIR is a technique to identify or recognize the image on the basis of features present in image. Basically features are classified in to four categories 1.color 2.Shape 3. texture 4. size .In this system we are extracting color, shape feature extraction. After that classifying images in to categories using neural network according to the weights and image displayed from the category for which neural network shows maximum weight. category1 belongs to wheat and category2 belongs to gram. Experiment was conducted on 200 images of wheat and gram by using Euclidean distance(ED) and artificial neural network techniques. From 200 images 150 are used for training purpose and 50 images are used for testing purpose. The precision rate of the system by using ED is 84.4 percent By using Artificial neural network precision rate is 95 percent.

Keywords: Features, Color, Shape, CBIR, Classification, ANN, Euclidean distance

1. INTRODUCTION

The application of machine vision is very important in agricultural industry. Seed analysis and classification can provide additional knowledge in their production, seeds quality control and in impurities identification. Generally these activities are performed by specialists by visually inspecting each sample, which is a very tedious and time consuming task [1]. So, automation is required in this field. Now a day, computer vision technology is applied in a large variety of fields to increase the efficiency of the work. So, This paper uses machine vision technique for the recognition aspect of the said problems[2].

In this paper a system is designed to recognize the different types of grains by their images on the basis of their features using content based image retrieval technique. Content-based image retrieval is a technique which uses visual contents to search images from large scale image databases according to users' interests. CBIR technique is further explained in next section. The proposed technique is based on color and shape features. And for classification two approaches are used and then compared. First classification is done through Artificial Neural Network(ANN) and second is done through finding the minimum Euclidean Distance between the features of two images.

2. CONTENT BASED IMAGE

RETRIEVAL

There are basically two types of image retrieval techniques 1.Text based image retrieval 2.Content based image retrieval. In the text based image retrieval technique, images are indexed on the basis of heading or topic, description, keyword. Texture based images can not be retrieved by text based query. To overcome this problem of text based image retrieval and reducing human effort in indexing process[3].

Content based image retrieval is a best technique to retrieve an image because it reduces the image indexing and texture based problems. efficiency of CBIR system can be improved by providing the feedback for a particular image which is not recognized by the system.

In a general CBIR system different types of features of image database is calculated and feature database is created. When the random test or input or query image is given to the system then all the defined features is extracted for that particular image and stored in feature vector.

Then image retrieval is done by comparing the similarity between the query feature vector and different feature vectors of feature database.

And then the image which has closest feature set as the query image is displayed as the result. The general architecture of CBIR system is shown in Fig 1[3].

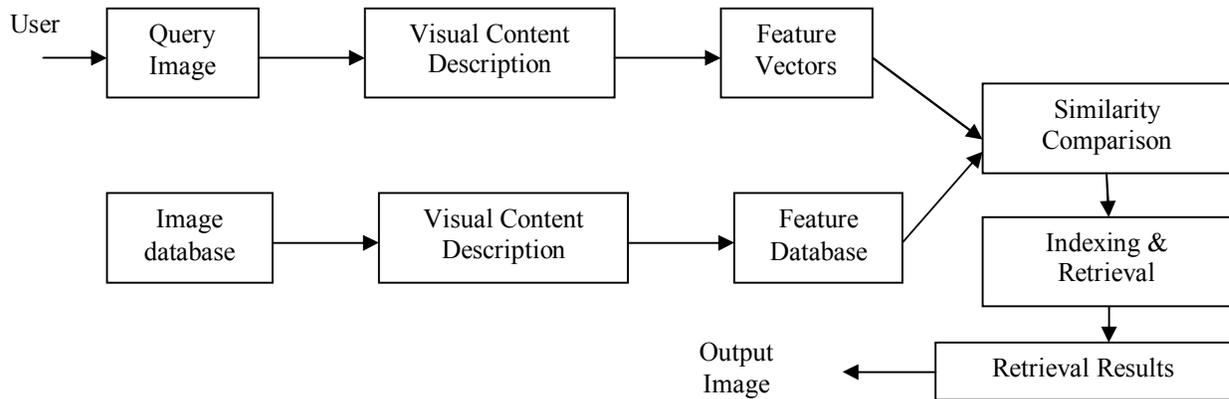


Figure 1. General architecture of CBIR

3. RELATED WORK

There are many researchers who develop a variety of seed image recognition system by applying different-different classification techniques e.g. a artificial neural network, a Euclidean distance technique, Histogram Intersection Distance etc. The details of each technique given below-

3.1 Euclidean Distance Method

Benjamaporn Lursthut and chomtip pornpanomchai developed a method which uses shape ,size ,color , texture feature extraction with the Euclidean distance technique to classify seed Images. The system's recognition rate was 95.1 % for trained dataset and 64.0 percent for unknown in untrained dataset [4].

Dr. H.B.Kekre, Mr. Dharendra Mishra, Ms. Stuti Narula and Ms. Vidhi Shah applied the color feature extraction to recognize different kinds of image. They applied Euclidean distance technique with The different images of the same class give results varied from 30% to 60% for a database of size 300 [5].

Poulami Haldar and Joydeep Mukherjee used Euclidean matrix method to calculate the distance vector. System provides overall accuracy 87.50 % for the images of different class [6].

3.2 Artificial Neural Network

Dayanand Savakar designed algorithms which is used to extract 18 color and 27 texture features from food grains. For the Recognition and Classification of Similar Looking

Food Grain Images this system uses artificial neural network. Recognition of Mustard is about 87% and for Soya is 78% using color feature and on the basis of texture feature extraction maximum classification rate is 84% [7].

Ai-Guo OuYang, Rong-jie Gao, Yan-de Liu,,Xu-dong Sun, Yuan-yuan Pan and Xiao-ling Dong designed a system in which color features in RGB and color space is computed. A back feed forward neural network trained to identify rice seed 86.5 % rice seeds were identified by the system [8].

3.3 Histogram Intersection Distance

Manimala Singha and K.Hemachandran used histogram intersection distance for feature similarity matching. Experiment performed on standard "Wang Database" containing 1000 image. In it texture and color feature extracted through wavelet transformation and color histogram [9].

3.4.Support vector machine

Ying Liua, Dengsheng Zhanga, Guojun Lua and Wei-Ying Mab used SVM in their system because of SVM has been used for object recognition, text classification, etc. After using SVM in that system an improvement of 10% in

retrieval accuracy is obtained compared with SVM (400 images for training) with much fewer training data [10].

b) Kantip Kiratiratanapruk and Wasin Sinthupinyo classify defects of corn seed in more than ten categories and extracted color, texture feature. They used SVM as type classifier. Accuracy of the system is 96.5 % for normal seed type and 86.5 % in the case of defect seed(group) types [1].

On the basis of above literature reviews it has been observed that ANN and Euclidean distance method of classification can provide better results with shape and color features.

4. PROPOSED METHODOLOGY

The general methodology of the proposed system is shown in Fig 2, which is further explained in subsections.

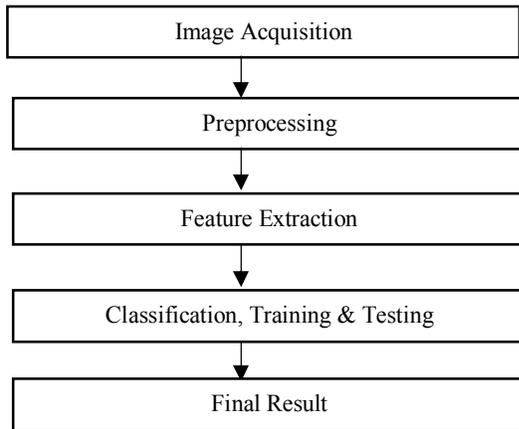


Figure 2. Algorithm of Proposed System

4.1 Image acquisition

12 mega pixel camera is used for taking the images of wheat, gram and pulse. Distance between seed and camera is almost equal to 14 centimeters.

4.2 Image preprocessing: Preprocessing operation includes the following steps-

4.2.1 Image resizing: The input images captured by different cameras may have different sizes which can affect the result, so initial resizing is necessary.

4.2.2 RGB to Gray Scale Conversion: Equation 1 is used to convert the RGB value of a pixel into its gray value.

$$\text{gray} = .2989 * R + .5870 * G + .1140 * B$$

4.2.3 Gray to Binary Image Conversion: Binarization is done using otsu method, it can be done using graythresh function in MATLAB.

4.2.4 Morphological Processing: Closing and Filling operations are performed using a disk type structuring elements of radius 2 to fill any holes in the images.

The steps of pre-processing on an input of wheat is shown in Fig 3.

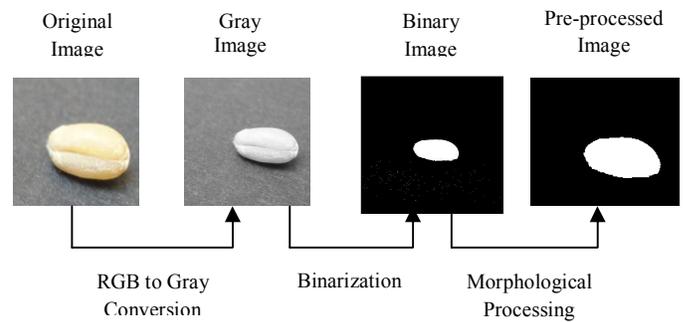


Figure 3. Pre-processing on an Input Image

4.3.Feature extraction

Following shape and color features are extracted for each Input Image.

4.3.1 Shape feature extraction: Following shape features are considered.

Seed roundness : Roundness of each seed is calculated w.r.t circle, means when the value of roundness of any particular seed is .9 then it's almost round. Roundness can be calculated by following equation.

$$R = 4 * \pi * \text{area} / P^2, \text{ where } P \text{ is the perimeter of the seed.}$$

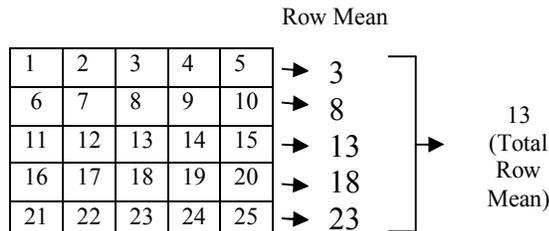
4.3.2 Color feature extraction

Row mean and column mean: For each Red, Green and Blue plane row and column mean is calculated by following step.

a) Three color planes Red, Green, Blue are separated and step 2 to 4 is performed on each plane.

b) For each plane row mean and column mean of colors are calculated.

Pictorially the row and column mean is calculated as follows [5]:



c) The row mean of all three planes are stored as 3 new features in feature vector.

Dominant color: Dominant Color is calculated for each Red, Green and Blue Plane by their histogram.

Histogram is representation of the occurrence of relative frequencies of various gray levels in an image. Dominant color is that gray level which has the maximum occurrence in the image. 3 more features are then added in the feature vector.

Median: It can be used to determine the value of intensity level of pixel which is separating high intensity value pixel from low intensity value pixel. Three median values are calculated for Red, Green and Blue Planes.

Some other statistical features are also calculated [11].

Standard Deviation: In the digital image processing it shows variation from standard or expected value. It is calculated for each plane by following function in MATLAB.

`std((std(Red,0,1)),0,1);`

Covariance: It is a positive number. It is the measure of change in two variable (random numbers) together.

`diag(cov(diag(cov(Red))))`

Kurtosis: It tells about the shape of probability distribution function of a random number. High kurtosis value is good for system because of it shows low noise and low resolution.

`kurtosis((kurtosis(Red)))`

Skewness: The value of skewness can be positive, negative, zero or may be undefined. Negative skewness shows that more values lie to the right of the mean, positive skewness

means more values lie to the left of mean and zero skewness shows values are evenly distributed on both sides of mean.

`skewness((skewness(Red)))`

Moment: The central first moment is zero and second central moment is calculated by using a divisor of N instead of N-1.

N -> length of vector or number of rows in matrix.

`moment((moment(Red,3)),3)`

4.4 Image Recognition

Image recognition is done using both ANN and Euclidean distance method.

Euclidean Distance Method

In this method Euclidean distance is calculated between a feature vector of test image and the feature file of all sample images stored in the database. The Euclidean distance can be calculated by using equation:

$$ED = \sqrt{\sum_{j=1}^m (c_j - b_j)^2}$$

Where ED is the Euclidean distance ,

m is number of features,

c_j is the value of feature j stored in the predefined database,

b_j is the value of feature

Artificial neural network

A neural network, illustrated in Fig. 4, consists of units (neurons), arranged in layers, which convert an input vector into some output. Each unit takes an input, applies a (often nonlinear) function to it and then passes the output on to the next layer. Generally the networks are defined to be feed-forward: a unit feeds its output to all the units on the next layer, but there is no feedback to the previous layer. Weightings are applied to the signals passing from one unit to another, and it is these weightings which are tuned in the training phase to adapt a neural network to the particular problem at hand. This is the learning phase.

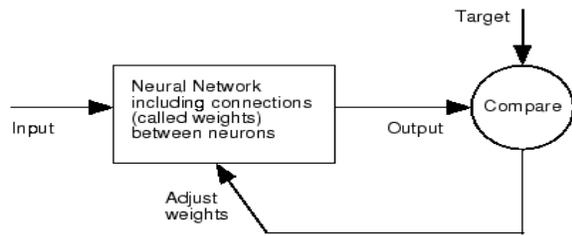


Figure 4. A typical Neural Network

Then the learned network with new weights is used for testing purpose.

5. EXPERIMENTAL RESULTS

The system is developed using MATLAB (R2011a), in this system 200 samples images of three different grains are taken by camera, from which 150 images are used for training phase and 50 for testing phase. Then the total 25 features of 150 sample images are calculated stored in a single feature file "feature_file.mat" using Matlab, this mat file stores features in a 25x150 array.

In classification using ANN method, first training of network is done using the previously generated "feature_file.mat" of 150 images, with 10 hidden neurons. Final application gives an option to user to select a test image form the rest 50 testing images and then the 25 features of this test image is calculated and stored in a 'test.mat' file, which is a 25x1 array. This "test.mat" file is then passed to the trained network. It gives 95% accuracy.

Euclidean Distance methods outputs that image from training set which has the closest features of test image. It gives 84.4 % accuracy.

Below Table 1 shows the values of different features for one test image of wheat.

Table 1. 25 Different feature values of Test Image

Roundness	.605
Red_row_mean	135.6
Green_row_mean	135.5
Blue_row_mean	133.5
Red_Dominatnt_color	144
Green_Dominat_color	144
Blue_Dominat_color	142
Red_Median	139
Green_Median	139
Blue_Median	137

Red_Std_Deviation	13.5
Green_Std_Deviation	12.0
Blue_Std_Deviaion	8.6
Red_Covariance	527920
Green_Covariance	380240
Blue_Covariance	144390
Red_Kurtosis	10.15
Green_Kurtosis	6.21
Blue_Kurtosis	1.97
Red_Skewness	1.05
Green_Skewness	1.00
Blue_Skewness	1.7
Red_Moment	57808420356433.4
Green_Moment	19499436210162.3
Blue_Moment	833492646434.140

And Fig 5. shows the final result, which has the test image and the closest matched image.

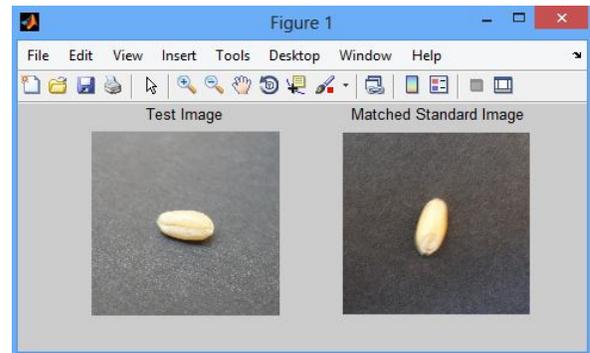


Figure 5. Final result of application with matched image

6. CONCLUSION

In the proposed system 25 shape, color and statistical features are calculated. Image recognition is done using both techniques Euclidean distance and artificial neural network. System is 95 % accurate using ANN and 84.4 % accurate using Euclidean distance method.

7. ACKNOWLEDGEMENTS

We would like to thank our guide Ms. Shanu Sharma for their guidance and feedback during the course of the project. We would also like to thank our department for giving us the resources and the freedom to pursue this project.

8. REFERENCES

- [1] Kantip Kiratiratanapruk and Wasin Sinthupinyo. 2011. Color And Texture For Corn Seed Classification By Machine Vision, International Symposium on Intelligent Signal Processing & Communication Systems(ISPACS). pp 1-5.
- [2] Adjemout Ouiza, Hammouche Kamal and Diaf Moussa. 2007. Automatic seed recognition by size, form and texture features, International Symposium on Signal Processing and its applications,(ISSPA). pp 1-4.
- [3] (2013),"Tutorial on CBIR", [Online] Available: www.cs.bgu.ac.il/~icbv061/...2006.../CBIR_Presentation.ppt
- [4] Benjamaporn Lursthut ,Chomtip Pornpanomchai, "Plant Seed Image Recognition System(PSIRS)", IACSIT International Journal of Engineering & Technology, 2011. pp. 600-605.
- [5] Dr. H.B.Kekre ,MR.Dhirendra Mishra, MS. Stuti narula and MS. Vidhi Shah , " Color Feature Extraction For Cbir", International Journal of Engineering Science & Technology(IJEST), 2011. pp. 8357-8365.
- [6] Poulami Haldar and Joydeep Mukherjee, "Content based Image Retrieval using Histogram,Color and Edge", International Journal of Computer Applications, 2012. pp 25-31.
- [7] Dayanand Savakar , "Recognition and Classification of Similar Looking Food Grain Images using Artificial Neural Networks", Journal of Applied Computer Science & Mathematics, 2012. pp 61-65.
- [8] Ai-Guo OuYang, Rong-jie Gao, Yan-de Liu,Xu-dong Sun, Yuan-yuan Pan. 2010. An Automatic Method For Identifying Different Variety Of Rice Seeds Using Machine Vision Technology, Sixth International Conference on Natural Computation. pp 84-88.
- [9] Manimala Singha and K. Hemachandran, "Content Based Image Retrieval using Color and Texture", Signal & Image processing: An International Journal, 2012. pp 39-57.
- [10] Ying Liua, Dengsheng Zhanga, Guojun Lua and Wei-Ying Mab. 2007. A survey of content-based image retrieval with high-level semantics", Pattern Recognition, Elsevier. pp 262-282.
- [11] Vijay Kumar, Priyanka Gupta, "Importance of Statistical Measures in Digital Image Processing", International Journal of Emerging Technology and Advanced Engineering, 2012. pp. 56-62.