International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 8, No.6 (Nov-2019)
http://dx.doi.org/10.12785/ijcds/080604

A Symbolic Representation and Classification of Fruits

Ranjit K N[1], Raghunandan K S[2], Naveen C[3], Chethan H K[2] and Sunil C[2]

[1] Department of Computer Science& Engineering, HKBK College of Engineering, Bengaluru, Karnataka, India.
[2]Department of Studies in Computer Science, University of Mysore, Mysuru, Karnataka, India.
[3]Department of Computer Science& Engineering, SJB Institute of Technology, Bengaluru, Karnataka, India.

Received 18 Mar. 2019, Revised 02 May 2019, Accepted 20 Oct. 2019, Published 1 Nov. 2019

Abstract: In the real word scenario, automation of digital system plays a vital role, especially in the field of agriculture which needs a good automated system for classification of different fruits as it consumes customer's time and is also very useful to the farmers. In this paper, we propose a system to classify different fruit classes using symbolic representation and classifier. Firstly, texture, color and shape features are extracted, and then natural clustering is applied on the feature fusion matrix. By making use of interval of mean and standard deviation, intra class variation is captured. For experimentation, 1200 images of 10 fruit classes is collected and totally 12000 samples are used. Further, symbolic classifier is used to obtain the confusion matrix and comparative study is made to show the robustness of symbolic representation and classifier with existing methods, SVM and KNN classifiers.

Keywords: Fruit Classification, Symbolic classifier, Clustering, Shape features.

1. INTRODUCTION

In today's world, India is the second largest country that holds agricultural land which yields 44.04 million tons of fruits and consists of 3.72 million hectares of land. Production of these fruits has been accounted by India around 10% across the world. Grapes, orange, mango, banana, papaya and apple are the vital fruits grown across India and has vast exports in fruits and horticulture with low cost production. Among all the industries, the fruit industry is considered as a crucial one as it contributes around 20% of the nation's growth. But from recent years we can observe that there is downfall in production of standard fruits because of inappropriate method of cultivation, maintenance and effective inspections. External quality of fruits plays a vital role and it has been assessed by their color, texture, shape and visual defects. It is also considered for various purposes such as export, producing juices etc. So this creates a challenging task for researchers to identify and detect the defects in the fruits at early stages, so that it doesn't spread to other fruits and prevents damaging / lowering the quality of fruits and as well as preventing economic losses to the country[1]

Some of the major attention and concern have been taken towards quality and safety of fruits from past few years. Technologies such as automation and intelligent sensing have the capability of restructuring our entire fruit processing and production units. Most of the data available in the agriculture applications are in the form of photographic images and to evaluate these information it is very difficult and possesses / poses some challenging task. Due the rapid growth in the field of information and science, Digital image processing and most recently computer vision based techniques are used for detailed analysis [2]. Some of the challenges connected with the identification and classification of fruit image samples of different classes of fruit images are shown in Fig.1.

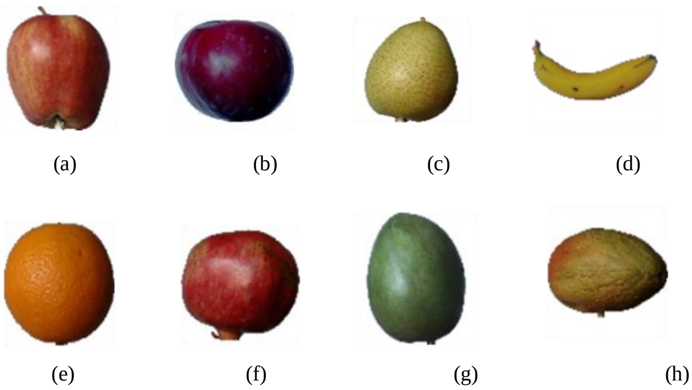

Figure. 1 Samples of Fruit Images (a) Apple (b) Plum (c) Pear (d) Banana (e) Orange (f) Pomegranate (g) Mango (h) Papaya.

E-mail: 22ranjukn@gmail.com, raghu0770@gmail.com,naveen.cse@gmail.com, hkchethan@gmail.com, sunilchaluvaiah87@gmail.com

The organization of proposed work is as follows, Section 2 gives a brief explanation on related work and Section 3 discusses the proposed model. Section 4 illustrates experimentation and results and Section 5 gives conclusion and future scope.

2. RELATED WORKS

Israr et. al., [3] proposed a novel technique for recognition of fruits which is built using Deep Convolution Neural Network by creating their own datasets that consist of 15 various classes of fruit images. Two ways of representation is adopted in this work, one to increase the system ability deep convolution neural network accompanied by max pooling and another one is incorporated by fully connected layers. Finally classification is done through probability procedures. Zaw Min Khaing et. al., [4] proposed a novel technique for developing a control system for recognition and classification of fruit images using Convolutional Neural Networks. For simulation, they have adopted the graphical processing unit which is available in the Matlab and for classification they used Alexnet and achieved good accuracy. Yu-Dong Zhang et. al., [5] proposed a novel method for classification of fruits based on deep convolution neural networks adapting 13 layers. It incorporates noise injection, gamma corrections and image rotation as their data augmentation methods and max pooling is differentiated with average pooling. CNN gives good results when compared to other state-of-art methods. Dahua Li et. al., [6] proposed a novel method for recognition of green apple fruit by combining the color, shape and texture features. For detection of green apple fruit shape and texture features are integrated and this method adopts HSV color space. Finally Support Vector Machine is adopted for the recognition of green apple fruit. Susovan Jana et. al., [7] proposed a novel method for recognition and classification of fruits based on shape which consists of seven classes. For Classification of fruits they have adopted various classifiers such as K-nearest neighbor, Naive Bayes and Neural network classifiers for each fruit class. Ranjit et. al., [8] proposed a novel technique for identification and classification of ten classes of fruit images which is associated with diseases. The method combines k-means and c-means algorithm for segmentation purpose and then for extracting features they have used Gray-level Co-occurrence matrix. For classification k nearest neighbor is applied and produces good results. Megha P Arakeri et. al., [9] proposed a novel technique for automatic grading of tomato fruit images by applying computer vision methods. It consists of two ways one is hardware and another is software development process, during hardware development, the images of fruit is captured and placed to their suitable bins and in software process fruit images are classified as defective and non-defective and ripe and unripe and gives effective results. Saswati Naskar et. al., [10] proposed a novel method for identification and classification of fruit images by making use of multiple features and Artificial Neural Networks. Texture, Shape and color features are used, whereas for texture feature Log Gabor filter is used and for color and shape mean hue has been used. Classification of fruit images is done by ANN and produce good results. Yudong Zhang et. al., [11] proposed a novel technique for classification of fruit images which uses fitness scaled chaotic artificial bee colony and feedforward Neural Network algorithm and also uses stratified k-fold cross validation and shows better results when compared other existing methods. Hossam M Zawbaa et. al., [12] proposed a method for classification based on random forest technique for three classes of fruit images. They have used three steps for this classification system pre-processing, feature extraction and classification and compared with k-NN and SVM classifier and gives better results. . Yudong Zhang et. al., [13] proposed a novel technique for classification of fruit images by making use of multi-class kernel support vector machine and after extraction of features, principal component analysis is adopted for decreasing the dimensions of feature space. For classification Multi-class SVM Contains, Max-Win-voting SVM, Directed Acyclic Graph SVM and Winner-Takes-All SVM and achieved good accuracy. S Arivazhagan et. al., [14] proposed a novel method for fruit recognition by integrating color and texture features. Classification of fruits is done based on Minimum Distance Criterion and achieved good recognition rate. Woo Chaw Seng et. al., [15] proposed a novel method for identification, classification and recognition of fruit images based on shape, color, and size. For classification of fruit images they have used k-nearest neighbors classifiers. In this method the fruit image are classified and recognized based on feature values obtained by KNN classification and achieved good accuracy. Jagadeesh Devdas Pujari et. al., [16] proposed a novel method for grading and classification of Anthracnose fungal disease of fruit images by using statistical texture features. In this method the affected lesion area of Anthracnose fruit images are separated from normal area using region growing, thresholding, marker controlled watershed and k-means clustering segmentation techniques. For classification of neural network classifier is used to identify affected Anthracnose fruit images from normal one and achieved good results. Esraa Elhariri et. al., [17] proposed a novel method for classification of Tomato ripeness based on Multi-class Support Vector Machine. This method consists of three stages, one is pre-processing, second is feature extraction using principal component analysis and classification using SVM classifier and achieved good classification accuracy. I Kavdir et. al., [18] proposed a method for classification of defective apple from good apple and also comparison of artificial neural networks and statistical classifiers using textual features for sorting apple. Apple fruit images which are extracted from Empire and Golden Delicious are classified using back propagation network with additive use of histogram features and statistical classifiers such as k-nearest neighbors, Decision Tree and Bayesian with textual features and proves statistical features yields better results. M A Shahin et. al., [19] proposed a novel method for classification of apple fruit images based on surface bruises by making use of neural network and image processing techniques. Spatial edge features and discrete cosine transform is used to indicate old bruises apple whereas artificial neural network is used for old bruises and new bruises apple and ANN classification system proves to be a robust method. R

Anand et. al., [20] proposed a novel method for classification of Multi-class problems based on Modular Neural Networks. In this work, modular network architecture is used for reducing the k-class problem to k-two class problems. This overcomes the problem of back propagation algorithm which is too slow for convergence of such multi-class problems. A S Simoes et. al., [21] proposed a novel method for automatic sorting of visual fruit using artificial neural networks. In this work, color based classification of orange fruit images using artificial neural network, a multi-layer perceptron along with back propagation algorithm is used and it proves to be robust to color variations for the proposed method. M A Shahin et. al., [22] proposed a new method for classification of apples based on watercore using artificial intelligence classifiers. In this work, the sorting of apples uses optimal neural network and fuzzy classifiers by making use of image feature as input variables. Artificial intelligence classifiers are compared with Bayesian classifiers and results in proving that neural classifier performs better.

3. METHODLOGY

In this work, we propose a novel method for classification of fruit images such as, Apple, Mango, Plum, Pear, Papaya, Banana and Orange, Pomegranate, Citrus limetta and Sapota. Different steps elaborated during the process of the proposed fruit classification model are shown in Fig. 2 and further explanations are given in sub sections below.

A. *Texture and Color Feature Extraction*

For feature extraction we make use of three feature extraction technique. Edge co-occurrence matrix (ECM) which is based on second order statistics of edge direction in an image. Two other feature extraction methods are Local binary patterns and Gabor filter features based on textures.

*a). Edge co-occurrence matrix (ECM).*It is almost similar to gray level co-occurrence matrix(GLCM), but the major difference is that GLCM make use of gray level image as their input were as, ECM make use of edge image as their input image[23]. During the process of ECM edge image are obtained from original gray images and these edges are detected in 8 directions and the strongest direction of the edge will be used as pixel location. Thus edge co-occurrence matrix is formed from

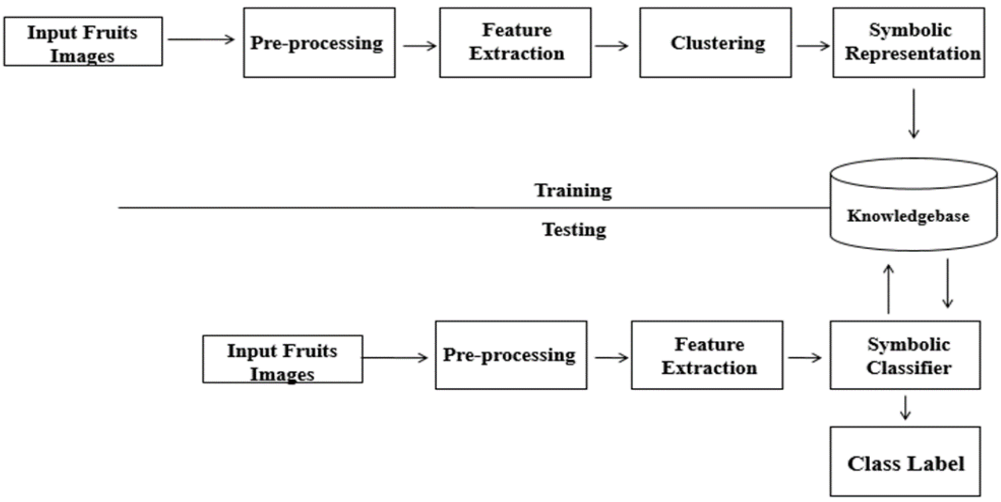

Figure 2. Block diagram of proposed fruit classification model

the pairs of edge pixels separated by a given displacement. Definition, of Edge co-occurrence matrix H, the $(i,j)^{\text{th}}$ element which is the number of appearances of edge direction $i$and j in the edge image

I is the distance and direction determined by the displacement vector d from each other is given by,

$$H_{ij} = \#\{x \mid \mathrm{I(x) = i\,, I(x+d) = j} \qquad (1)$$

I is the edge image, $d = (d_x, d_y)$ is the displacement vector, $\neq$ is the number of elements in the set and

$X = (x, y)$ runs through the edge image$I$. Since edges were detected in 8 directions, the size of ECM is 8 x 8.

b). Local Binary Pattern. The normal idea of Local Binary Pattern (LBP), is 2-Dimensional surface textures that can demonstrate two kinds of measure such as, gray scale contrast and local spatial patterns [24]. The definition of $LBP_{P,R}$ given as,

$$LBP_{P,R} = \sum_{p=0}^{p-1} S(g_p - g_c)2^p \qquad (2)$$

Where,

$$S(x) = \begin{cases} 1, if\ x >= 0 \\ 0, otherwise \end{cases}$$

From the above equation $S(x)$ denotes thresholding function, g_c and g_p are the gray levels of center pixel, around the radius R, P is the adjacent pixels in circular neighborhood.

c). Gabor Filter. This is local texture descriptor that is named after Dennis Gabor. Frequency characteristics and spatial orientation represents Gabor wavelet which relates to human visual system and also which is identical to Windowed Fourier transform [25]. It is used in numerous applications because it acts as multi-resolution filter. To extract local features 2D wavelet decomposition and Gabor filter is used. In spatial domain, 2D Gabor filter is modulated by Gaussian Kernel function and sinusoidal plane wave. Implementing Gabor filter we can procure real and imaginary features such as frequency λ and orientations θ given as,

$$\lambda(x, y, f, \theta, \sigma_x, \sigma_y,) = \frac{1}{2\pi\sigma_x\,\sigma_y} exp\left\{-\frac{x^2+y^2}{2\sigma^2} + \lambda(x\cos\theta + y\sin\theta)\right. \quad (3)$$

B. Shape Features Extraction

Basically simple geometric features can be used to relate shapes. Normally the geometric features are used to differentiate shapes with huge differences and to eliminate false hits. It is used as filters or combined with other shape descriptors for differentiating shapes. Some of the shape parameters are Center of Gravity, Axis of least inertia, Average bending energy, Circularity Ratio, Rectangularity, Convexity, Solidity, Euler Number, Profiles and Hole area ratio that are explained in the sub sections below.

a). Center of Gravity

Center of Gravity is also known as Centroid and its position is fixed based on its shape. Shape which is represented by its region and contour is given by equation (4) and (5).

$$\begin{cases} g_x = \frac{1}{N}\sum_{i=1}^{N} x_i \\ g_y = \frac{1}{N}\sum_{i=1}^{N} y_i \end{cases} \quad (4)$$

Where,
N = number of point in shape
$(x_i, y_i) \in \{(x_i, y_i) | f(x_i, y_i) = 1\}$.

$$\begin{cases} g_x = \frac{1}{6A}\sum_{i=0}^{N-1}(x_i + x_{i+1})(x_i\,y_{i+1} - x_{i+1}y_i) \\ g_y = \frac{1}{6A}\sum_{i=0}^{N-1}(y_i + y_{i+1})(x_i\,y_{i+1} - x_{i+1}y_i) \end{cases} \quad (5)$$

Where A is the area of the contour and is given by,

$$A = \frac{1}{2}|\sum_{i=0}^{N-1}(x_i\,y_{i+1} - x_{i+1}y_i) \quad (6)$$

b). Axis of least inertia

Definition of Axis of least inertia (ALI), is the line for which the integral of the square of the distances to the points on the shape boundary is minimum. If we consider x $\sin\theta - y\cos\theta = 0$ as the parameter equation of ALI, then the slope angle θ is given by, and α is the angle between the axis of least inertia and x-axis. Inertia is given by equation (5) and (6).

$$I = \frac{1}{2}(a + c) - \frac{1}{2}(a - c)\cos 2\alpha - \frac{1}{2}b\sin 2\alpha \quad (7)$$

Where,
$a = \sum_{i=0}^{N-1} x_i^2$, b = $2\sum_{i=0}^{N-1} x_i\,y_i$, $c = \sum_{i=0}^{N-1} y_i^2$

c). Average bending energy

Definition of Average bending energy [27] is given by,

$$BE = \frac{1}{N}\sum_{s=0}^{N-1} Ks^2 \quad (8)$$

where,
$k(s)$, curvature function
s, arc length parameter
N, number of points on the contour

d). Circularity Ratio

Circularity Ratio is defined as the ratio of the area of the shape to the area of the circle with the same perimeter and it is given as [28],

$$C_r = \frac{A_s}{A_c} \quad (9)$$

Where,
A_s= area of the shape,
A_c = area of the circle possessing the same perimeter of the shape.

e). *Rectangularity*

The meaning of rectangularity is that the shape of its rectangular is filled to its minimum bounding rectangle and it is given as,

$$\text{Rectangularity} = \frac{A_S}{A_R} \quad (10)$$

Where,
A_S = area of the shape,
A_R = area of minimum bounding rectangle.

f). Convexity

Convexity can be defined as, ratio of perimeters of convex hull and original contour [29] and is given as,

$$Convexity = \frac{O_{convexhull}}{O} \quad (11)$$

Where, O = original contour

g). Solidity

Solidity illustrates the range of the convex or concave in shape and solidity convex contour is consistently 1, and the definition is given as [30],

$$\text{Solidity} = \frac{A_s}{H} \quad (12)$$

Where,
A_s= area of the shape region,
H = convex hull area of the shape.

h). Euler Number

It illustrates the correlation between the number of contiguous parts and the number of holes on the shape and we consider S as the number of contiguous parts and N as the number of holes in the shape and is given as,

$$Eul = S - N \quad (13)$$

i). Profiles

Projection of the shape to x-axis and y-axis to their Cartesian coordinate system and produce two one dimension functions,

$$Pro_x(i) = \sum_{j=jmin}^{jmax} f(i,j) \text{ and } Pro_x(j) = \sum_{i=imin}^{imax} f(i,j) \quad (14)$$

Where, $f(i,j)$is region of shape.

j). Hole area ratio

Hole Area Ratio(HAR) is the most efficient in perception between symbols with big holes and symbols with small holes [31]and definition of HAR is given as,

$$HAR = \frac{A_h}{A_s} \quad (15)$$

Where,
A_h = area of the shape,
A_s = total area of all holes in the shape.

C. One Dimensional Function for Shape Representation

Normally one dimensional function is obtained from the coordinates of the boundary of shape and this is known as shape signature [32,33] and it captures perceptual feature of the shape [34].Some the Shape Signature used are Curvature function, Area of function, Complex coordinates, Tangent angle, Triangle area representation, Centroid distance function and Chord length function.

a). Complex coordinates

Complex number that are initiated from the coordinates of the boundary points are known as Complex Coordinates$P_n(x(n), y(n)), n\epsilon[1, N]$,

$$z(n) = [x(n) - g_x] + i[y(n) - g_y \quad (16)$$

Where,
(g_x, g_y) = Centroid of the shape.

b). Centroid Distance Function

The distance of the boundary points from the Centroid (g_x, g_y) of the shape is called as Centroid distance function,

$$r(n) = [(x(n) - g_x)^2 + (y(n) - g_y)^2]^{1/2} \quad (17)$$

Subtraction of Centroid constitutes to position of shape, from the boundary coordinates, hence centroid distance and complex coordinates will be invariant to translation.

c). Tangent angle

The tangent angle function at a point $P_n\,(x(n), y(n))$will be defined as tangential direction of contour [35].

$$\theta(n) = \theta_n = \arctan\frac{y(n)-y(n-w)}{x(n)-x(n-w)} \quad (18)$$

and the every contour is digital curve, w is a small window to calculate $\theta(n)$ more accurately.

d). Contour curvature

Identification of resemblance between the shapes from human is very important; hence Curvature acts as a vital boundary feature and is a key perceptual characteristic. It produces effective results for shape recognition [36]. The function of curvature $K(n)$ is taken from [37, 38],

$$K(n) = \frac{\dot{x}(n)\ddot{y}(n) - \dot{y}(n)\ddot{x}(n)}{(\dot{x}(n)^2 + \dot{y}(n)^2)^{3/2}} \quad (19)$$

f). Area of Function

When there is a change in boundary points, the shape boundary also changes, the area of the triangle will be created from two consecutive boundary points and center of the gravity will change. Hence this kind of area function can be utilized as shape representation. The two consecutive boundary points are P_n ,$P_{n+1}$and the center of gravity is given as, G.

D. Moments

a). Boundary Moments

For minimizing the dimensions of boundary representation we can make use of analysis of contour and Boundary moments [39] and we consider shape boundary presented in the form of 1-D shape representation $z(i)$ and the r^{th} moment m_r and the central moment μ_r is given as,

$$\mu_r = \frac{1}{N}\sum_{i=1}^{N}[z(i)]^r \text{ and } \mu_r = \frac{1}{N}\sum_{i=1}^{N}[z(i) - m_1]^r \quad (20)$$

where, N= number of boundary points.

E. Region Moments

In all forms of region based descriptors, the most popular one is the moments and it also incorporates Radial chebyshevmoments, Zernike moments and invariant moments. The normal idea of moments function m_{pq} is in the order of $(p + q)$ of the shape region and it is shown as,

$$m_{pq} = \sum_x \sum_y \Psi_{pq}(x,y) f(x,y) \text{p, q} = 0,1,2,3,.... \quad (21)$$

a). Invariant Moments (IM)

Invariant moments is also known as geometric moment invariant and these Geometric moments are effortless of the moment function with basis $\Psi_{pq} = x^p y^p$ which is complete and not orthogonal [40]. Geometric moments m_{pq} is in the order of $(p + q)$ and it is given as,

$$m_{pq} = \sum_x \sum_y x^p y^p f(x,y) \text{p, q} = 0,1,2,3 \ldots. \quad (22)$$

b). Algebraic moments Invariant

From the first m central moments the algebraic moment invariant will be calculated and then it is given in the form of eigenvalues of predefined matrices $M_{[i,j]}$whereas the elements are the scaled factors for the central moments. Hence the algebraic moment invariants are invariant affine to the transformations and it is capable of constructing up to arbitrary order. Objects that are having various configuration outlines provides very good and very bad results when it is subjected to the algebraic moment invariants [41].

c). Zernike moments (ZM)

The Zernike moments is also known as orthogonal moments and orthogonal Zernike polynomials derives the complex Zernike moments [42].

$$V_{nm}(x,y) = V_{nm}(r \cos\theta, \sin\theta) = R_{nm}(r)\exp(jm\theta) \quad (23)$$

where, $R_{nm}(r)$ = orthogonal radial polynomial

d). Radial Chebyshevmoments (RCM)

The basic definition of RadialChebyshev moments is of the order p and repetition q is given in the form of [43],

$$S_{pq} = \frac{1}{2\pi\rho(p,m)} \sum_{r=0}^{m-1} \sum_{\theta=0}^{2\pi} t_p(r).\exp(-jq\theta).f(r,\theta) \quad (24)$$

where, $t_p(r)$ = scaled orthogonal Chebyshevpolynomials.

F. Clustering and Symbolic Representation

Here, related fruit images are grouped together to their own categories. These images will be clustered by applying partition clustering which is very easy for the adopted feature matrix to cluster preferably proximity matrix as in hierarchal clustering. K-means clustering method is adopted for clustering identical fruit images. To capture inter-class variations we have applied interval data representation for clustered fruit images [26].

By considering a sample $z_i = \{w^1, w^2, w^3, \ldots\ldots w^d\}(z_i \epsilon u^d)$ that lies in the i^{th} class including d features. Hence, total number of S sample sp is the number of classes. Thus, after applying clustering to the samples that lies in the i^{th} class and number of clusters acquired from each class k and the total number of samples of j^{th} cluster that belongs to class $i$is $s_j^i j = 1,2,\ldots,kandi = 1,2,\ldots,p$. To capture the intra-class differences from each cluster Mean-standard deviation interval representation is used and the mean and standard deviation evaluated for the clustered samples is specified in equation (25) and (26).

$$\mu_{j_i}^l = \frac{1}{s_j^i} \sum_{h=1}^{s_j^i} w_h^l \quad (25)$$

$$\sigma_{j_i}^l = \sqrt{\frac{1}{(s_j^i - 1)} \sum_{h=1}^{s_j^i} \left(w_h^l - \mu_{j_i}^l\right)} \quad (26)$$

Where, $\mu_{j_i}^l$ and $\sigma_{j_i}^l$ are the mean and standard deviation value of l^{th} feature that belongs to j^{th} cluster corresponding to class i respectively. Further the mean and standard deviation is computed for all features belongs to j^{th} cluster corresponding to i^{th} class.

Once the mean and standard deviation are computed for each cluster that belongs to a particular class, then these two moments will be joined together to form an interval cluster representative that belongs to each class. Then, the difference between mean and standard deviation represents the lower limit of an interval and the sum of mean and standard deviation represents the upper limit of an interval. Eventually, k number of such cluster interval representatives are obtained from each class. Cluster representatives is given by,

$$CS_j^i = \{[(\mu_{j_i}^1 - \sigma_{j_i}^1), (\mu_{j_i}^1 + \sigma_{j_i}^1)], [(\mu_{j_i}^2 - \sigma_{j_i}^2), (\mu_{j_i}^2 + \sigma_{j_i}^2)], \ldots, [(\mu_{j_i}^b - \sigma_{j_i}^b), (\mu_{j_i}^b + \sigma_{j_i}^b)]\}$$

$$CS_j^i = \{[f_1^-, f_1^+], [f_2^-, f_2^+], \ldots, [f_b^-, f_b^+]\}$$

where, $f_l^- = \{(\mu_{j_i}^l - \sigma_{j_i}^l)\}$ and $f_l^+ = \{(\mu_{j_i}^l + \sigma_{j_i}^l)\}$

Ultimately, we arrived at an interval feature matrix of dimensions (k*q) zb, which is considered as a reference matrix for further classification.

G. Fruit Classification

We adopted symbolic classifier in our proposed classification method to check the efficacy for classification of scripts in video frames. In this, the reference script images are represented in the form of interval data as explained above. We consider a test sample$V_t = \{v^1, v^2, \ldots, v^b\}$, that contains b number of features and the test sample v_t should be classified into any one of the four member classes. We have computed similarity between the test samples and all the reference samples and for each test sample similarity is computed at feature level.

Therefore, the similarity between a test crisp (single valued) feature and reference interval feature is computed as shown below: When the crisp value lies between the upper limit and lower limit, then the value of similarity is 1 or else 0. Identically, the similarity betweenU_t and rest of the samples are computed. If U_t is said to be a member of any one of the four classes, then the value of acceptance count $AC_q^{j_i}$ is very high with respect to the reference sample (cluster representative) that belongs to a particular class. Then, the acceptance count $AC_q^{j_i}$ for a test sample corresponding to j^{th} cluster of i^{th} class is given by:

$$AC_q^{j_i} = \sum_{l=1}^{b} \text{Sim}\left(V_t, CS_j^i\right) \quad (27)$$

where,

$$\mathrm{Sim}\left(V_t\,, CS_j^i\right) = \begin{cases} 1 & \text{if} v^l \geq f_l^- \; and \; v^l \leq f_l^+ \; a \\ & \quad and \\ 0 & \quad \text{otherwise} \end{cases}$$

4. EXPERIMENTATION AND RESULTS

In this section we describe the experimental setup and comparative study on the proposed symbolic classifier and conventional classifiers. To evaluate the proposed method we have collected 1200 samples from each class. We have considered 10 fruit classes for evaluation and totally our database consists of 12000 sample images for classification. In our proposed classification system, the dataset is divided randomly into training and testing. Seven sets of experiments have been conducted under varying number of training set images as 20%, 30%, 40%, 50%, 60%, 70% and 80%. At each training stage, the fruit images are represented in the form of interval data with respect to the varied number of clusters from 6 to 15. Values less than 6 clusters do not give good results, so we ignore the results of cluster value less than 6. At testing stage, the system uses remaining 80%, 70%, 60%, 50%, 40%, 30%, and 20% of fruit images respectively for classifying them as any one of the three classes. The experimentation in testing is repeated for 20 different trials. During testing, the classification results are presented by the confusion matrix. The performance of the classification system is evaluated using classification accuracy, precision, recall, and F-Measure computed from the confusion matrix. The proposed method is implemented using MATLAB 2018Ra software. The classification results are thus obtained for different training and testing percentage of samples under varied cluster sizes from 6 to 15. These results are measured in terms of accuracy (minimum, maximum, and average), precision (minimum, maximum, and average), recall (minimum, maximum, and average) and F-Measure (minimum, maximum, and average). The minimum, maximum, and average of respective results are obtained from 20 trials of experiments performed on training samples. Here, precision and recall are computed from the results obtained from the class wise precision and class wise recall respectively. For every cluster value 20 trails are made to obtain the average confusion matrix. From Table IV we can observe that when the cluster value is at 12, it gives better results for classification. In the same way we compared the results with convention classifiers viz., Support Vector Machine (SVM) and K Nearest Neighbor (KNN). Our symbolic method gives good performance compared to other two conventional classifiers. The proposed symbolic classifier confusion matrix results is show in Table I and Table II and Table III shows the confusion matrix results of SVM and KNN classifiers respectively. In Table IV, the experimental results of Recall, Precision and F-measure for different training and testing sets are shown and also the cluster sizes are varied to obtain better results. In Table V, the comparative study on existing fruit classification algorithms is showed. Compared to the existing methods our methods outperform with better accuracy. Fig. 3., Fig. 4. and Fig. 5. shows how the k-values are decided to obtain final classification results.

The performance of the proposed symbolic method performance is better because it captures the intra class similarity. It helps to classify fruits with similar shape since we used shape features for classification. For classification using SVM we have adopted RBF kernel and to calculate the accuracy 10-fold cross validation has been used with various kernel parameters γ and cost parameters c for each binary classifier $\gamma = [2^{-15}, 2^{-14}, 2^{-13}, \ldots, 2^{15}]$ and $c = [2^{-15}, 2^{-14}, 2^{-13}, \ldots, 2^{15}]$. Whichever gives better accuracy among the parameters is adopted. In the same way for KNN we used K value 1 with 10 fold cross validation to obtain average confusion matrix. From the overall summary we conclude that the proposed symbolic method outperformed compared to other SVM and KNN conventional classifiers. Fig. 6. shows sample results of correct classification from the proposed method.

TABLE I. PROPOSED SYMOBOLIC CLASSIFIER CONFUSION MATRIX RESULTS OF AVERAGE OF 20 TRAILS.

Classes	Confusion Matrix from Symbolic classifier.									
Apple	**95.2**	0.3	2	1.1	1.4	0	0	0	0	0
Mango	1.6	**94.3**	0.5	0	0	1.2	0	1.8	0	0.6
Plum	0.8	1.3	**95.6**	0.4	0	0	0	0.8	1.1	0
Pear	0.6	0.8	0.2	**96.4**	0	0	0	0.9	0.8	0.3
Papaya	0	0	0	0	**95.8**	2.4	1.8	0	0	0
Banana	0	0.2	0.6	0	2.3	**93.7**	0	0	1.6	1.6
Orange	0	0	0	0	0	0	**96**	0	0	4
Pomegranate	0	3.7	0	0	0	0	0.8	**93.4**	1.4	0.7
Citrus limetta	2.1	3.4	5.7	0	0	0	0	0	**85.2**	3.6
Sapota	0	0	2.7	0	0	0	1.5	3.7	1.4	**90.7**

TABLE II. SVM CLASSIFIER CONFUSION MATRIX RESULTS OF AVERAGE OF 10 FOLD CROSS VALIDATION.

Classes	Confusion Matrix from SVM classifier.									
Apple	**62.2**	3.5	2.8	10.7	5.4	6.3	0	0	4.5	4.6
Mango	3.8	**70.4**	0	5.3	7.2	0	2.6	10.7	0	0
Plum	0	0	**71.6**	11.6	2.7	2.4	8.4	0	0	3.3
Pear	3.3	2.2	0	**68.7**	0	0	10.6	15.2	0	0
Papaya	0	0	0	12.5	**64.3**	0	0	10.3	6.2	6.7
Banana	4.4	0	5.7	0	6.4	**75.7**	0	0		7.8
Orange	5.5	0	6.2	0	7.6	0	**70.6**	0	5.8	4.3
Pomegranate	3.4	2.7	6.1	4.6	0	0	7.4	**62.3**	5.2	8.3
Citrus limetta	7.5	5.3	5.7	5.5	4.8	6.2	0	0	**60.5**	4.5
Sapota	2.1	5.3	2.7	5.7	0	0	2.4	3.4	0	**78.4**

TABLE III. KNN CLASSIFIER CONFUSION MATRIX RESULTS OF AVERAGE OF 10 FOLD CROSS VALIDATION.

Classes	Confusion Matrix from KNN classifier.									
Apple	**65.7**	3.1	6.7	8.7	4.2	5.6	0	0	2.3	3.7
Mango	3.5	**63.4**	0	6.2	5.8	0	3.7	5.5	6.5	5.4
Plum	0	6.4	**68.6**	7.3	5.2	4.5	2.6	2.2	0	3.2
Pear	2.6	3.7	0	**62.1**	2.8	3.6	7.3	12.4	3.7	2.8
Papaya	0	2.7	0	5.2	**68.4**	3.2	2.3	5.2	8.3	5.7
Banana	4.4	0	5.7	3	6.4	**70.3**	0	0	2.4	7.8
Orange	2.4	0	5.8	0	6.2	0	**78.2**	0	4.8	2.6
Pomegranate	2.1	2.7	4.5	3.7	0	0	3.6	**70.4**	5.8	7.2
Citrus limetta	6.4	4.7	3.5	4.2	3.8	6.4	0	0	**68.2**	2.8
Sapota	4	6.3	2.7	6.8	2.5	0	2.5	3.4	0	**71.8**

TABLE IV. BEST RESULTS OBTAINED FROM ALL CLUSTERS UNDER VARIED TRAINING AND TESTING PERCENTAGE OF SAMPLES.

Samples Train-Test %	Accuracy			Precision			Recall			F-Measusre			Cluster #
	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg	
20-80	75.32	78.54	76.93	70.12	74.15	72.14	62.57	68.14	65.36	66.13	71.02	68.58	15
30-70	76.54	78.73	77.64	71.94	76.32	74.13	65.84	70.43	68.14	68.75	73.26	71.01	14
40-60	76.73	82.47	79.6	71.56	80.41	75.99	70.58	77.32	75.95	71.07	78.83	75.97	14
50-50	79.26	85.68	82.47	73.64	85.38	79.51	74.26	81.68	78.97	73.95	83.49	79.24	14
60-40	81.67	88.27	84.97	78.77	90.23	84.5	76.38	85.94	82.16	77.56	88.03	83.31	13
70-30	88.46	94.81	91.64	82.53	92.64	85.59	76.55	87.53	82.04	79.43	90.01	83.78	12
80-20	**92.33**	**96.42**	**93.63**	**86.92**	**93.2**	**91.06**	**81.75**	**90.78**	**86.27**	**84.26**	**91.97**	**88.60**	12
Best	**92.33**	**96.42**	**94.38**	**86.92**	**93.20**	**90.06**	**81.75**	**90.78**	**86.27**	**84.26**	**91.97**	**88.12**	

TABLE V. CLASSIFICATION ACCURACY BASED ON DIFFERENT TRAINING ALGORITHMS.

Algorithms	Classification Accuracy
Ref. [45]	85.72
Ref. [44]	83.23
Ref. [47]	80.36
Ref. [46]	82.82
Ref. [48]	85.44
Ref. [49]	92.18
Proposed Method	**93.63**

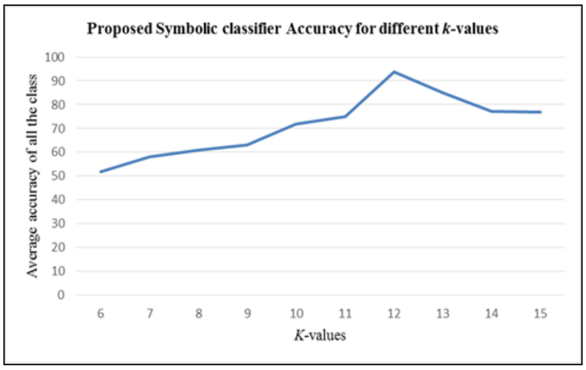

Figure 3. Accuracy of Proposed Symbolic Classifier for different k-values.

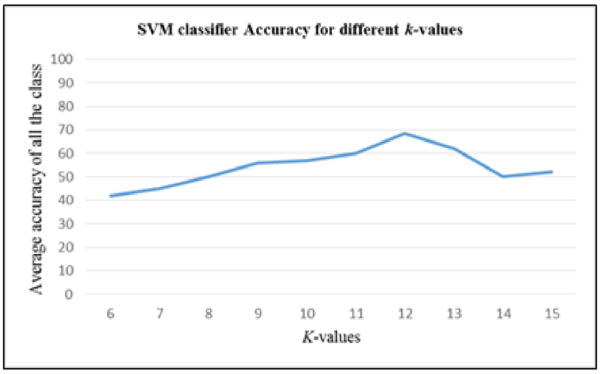

Figure. 4. Accuracy of SVM Classifier for different k-values.

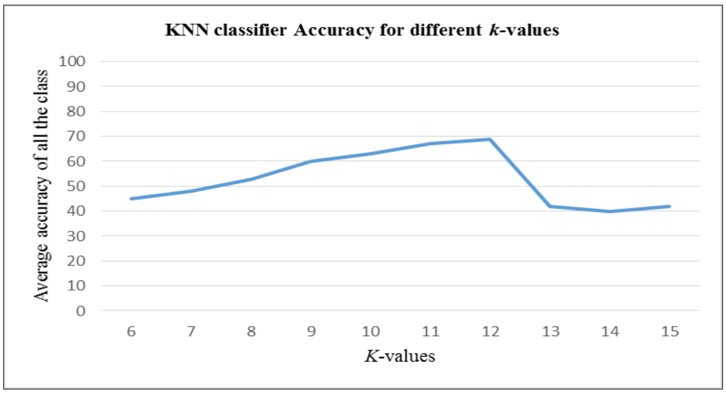

Figure. 5. Accuracy of KNN Classifier for different k-values

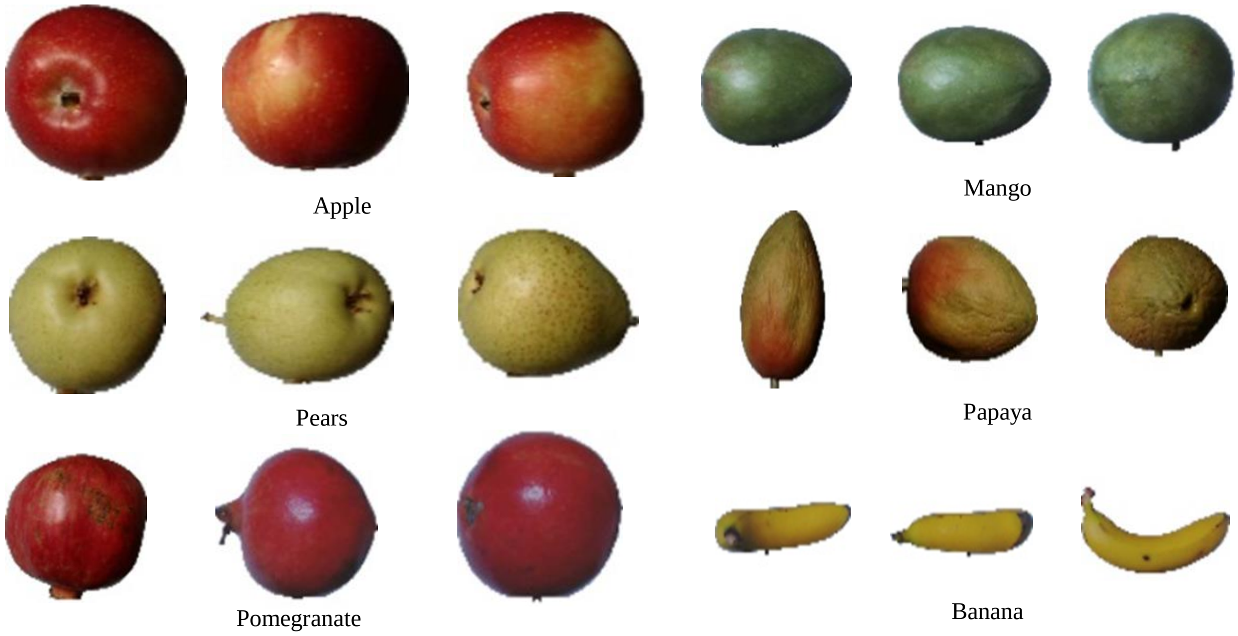

Figure 6. Some successfully classified images.

5. Conclusion and Future Scope

In this paper, a symbolic representation and classifier for fruit classification is proposed. The work considers 10 fruit classes for classification. Texture, color and also shape features are extracted and then natural clustering is used. The cluster values are varied from six to fifteen to obtain better classification results for different training and testing samples. For every cluster average of 20 trails are taken for the confusion matrix and cluster value 12 gives good results. Extensive experimentation is made and results are compared with existing methods and conventional classifiers viz. SVM and KNN. Confusion matrix and classification accuracy result shows that the proposed method gives better performance compared to conventional classifiers results. In future, we plan to use deep learning methods to improve the classification results.

References

[1]. J. D. Pujari R. Yakkundimath and A.S. Byadgi,: Grading and Classification of Anthracnose Fungal Disease of Fruits based on Statistical Texture Features, International Journal of Advanced Science and Technology Vol. 52, 2013.

[2]. A. Bhargava, A. Bansal,: Fruits and Vegetable Quality Evaluation Using Computer Vision: A review, Journal of King Saud University- Computer and Information Sciences, Elsevier, 2018.

[3]. I. Hussain, Q. He, Z. Chen,: Automatic Fruit Recognition Based on DCNN for Commercial Source Trace System, International Journal on Computational Science and Application (IJCSA), Vol 8, 2018.

[4]. Z. M. Khaing, Y. Naung, P. H. Htut,: Development of Control System for Fruit Classification Based on Convolution Neural Network, IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, 2018.

[5]. Y. D. Zhang, Z. Dong, X. Chen, W. Jia, S. Du, K. Muhammad, S. H. Wang,: Image based Fruit category Classification by 13-layer deep convolutional neural network and data augmentation, Multimed Tools Appl, Springer, 2017.

[6]. D. Li, M. Shen, D. Li, X. Yu,: Green Apple Recognition method based on the Combination of Texture and Shape Features, International Conference on Mechatronics and Automation (ICAM), IEEE, 2017.

[7]. S. Jana, R. Parekh,: Shaped-based Fruit Recognition and Classification, International Conference on Computational Intelligence, Communications, and Business Analytics, (CICBA), Springer, 2017.

[8]. K. N. Ranjit, H. K. Chethan, C. Naveena,: Identification and Classification of Fruit Diseases, International Journal of Engineering Research and Application(IJERA), 2016.

[9]. M. P. Arakeri, Lakshmana,: Computer Vision based Fruit Grading System for Quality Evaluation of Tomato in Agriculture Industry, International Conference on Communication, Computing and Virtualization, Elsevier, 2016.

[10]. S. Naskar, T. Bhattacharya,: A Fruit Recognition Technique using Multiple Features and Artificial Neural Networks, International Journal of Computer Applications, 2015.

[11]. Y. Zhang, S. Wang, G. Ji, P. Phillips,: Fruit Classification using Computer Vision and feedforward Neural Network, Journal of Food Engineering, Elsevier, (2014).

[12]. H. M. Zwabaa, M. Hazman, M. Abbass, A. E. Hassanien,: Automatic Fruit Classification Using Random Forest Algorithm, International Conference on Hybrid Intelligent Systems, IEEE, (2014).

[13]. Y. Zhang and L. Wu,: Classification of fruits using computer vision and a Multiclass Support Vector Machine, Sensors, 12(9), 12489–12505. doi:10.3390/s120912489 , 2012.

[14]. S Arivazhagan, R N. Shebiah, S. S. Nidhyanandhan, L Ganesan,: Fruit Recognition using Color and Texture Features, Journal of Emerging Trends in Computing and Information Sciences, 2010.

[15]. W. C. Seng, and S. H. Mirisaee,: A New Method for Fruit Recognition System, International Conference on Electrical Engineering and Informatics,Selangor, Malaysia, 2009.

[16]. J. D. Pujari, R. Yakkundimath and A.S. Byadgi,: Grading and Classification of Anthranose Fungal Disease Fruits based on Statistical Texture features, International Journal of Advanced Science and Technology, 2013.

[17]. E. Elhariri, N. El-Bendary, M. M. M. Fouad, J. Platos, A. E. Hassanien and A. M M. Hussien.: Multi-class SVM based Classification Approach for Tomato Ripeness, Innovations in Bio-inspired Computing and Applications, Springer, Cham 2014.

[18]. I Kavdir, and D E Guyer,: Comparison of Artificial Neural Network and Statistical Classifiers in Apple Sorting using Textual Features, Biosystems Engineering, Elsevier, 2004.

[19]. M A Shahin, E W Tollner, R W McClendon, H R Arabnia,: Apple classification based on Surface Bruises using Image Processing and Neural Networks, Transactions of ASAE,2002.

[20]. R Anand, K Mehrotra, C K Mohan, S Ranka,: Efficient Classification for multi-class problems using Modular Neural Networks, Transactions on Neural Network, IEEE, 1995.

[21]. A S Simoes, A H Reali Costa, A R Hirakawa, A M Saraiva,: Applying Neural Networks to Automated Visual Fruit Sorting, Proceedings of the World Congress of Computers in Agriculture and Neural Resources, 2002.

[22]. M A Shahin, E W Tollner, R W McClendon,: Artificial Intelligence Classifiers for sorting Apples based on Watercore, Journal for Agriculture Engineering Research, Elsevier,2001.

[23]. R. Rautkorpi and J. Iivarinen,: A Novel Shape Feature for Image Classification and Retrieval, International Conference on Image Analysis and Recognition, ICIAR, Springer, 2004.

[24]. T Ojala, M Pietikainen, T Maenpaa,: MultiresolutionGray-Scale Invariant Texture Classification with Local Binary Patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence. vol. 24, No.7, 971-981, 2002.

[25]. T. Ojala, M. Pietik¨ainen, D. Harwood.: A comparative study of texture measures with classification based on feature distributions. Pattern Recogn. 29(1), 51–59 1996.

[26]. D S Guru, N Vinay Kumar,: Symbolic Representation and Classification of Logos, Proceedings of International Conference on Computer Vision and Image Processing CVIP 2016.

[27]. I Young, J Walker, J Bowie,: An analysis Technique for biological shape. Information and Control, Elsevier, 1974.

[28]. D Zhang, Guojun Lu,: Review of Shape representation and description techniques, Pattern Recognition, Elsevier, 2004.

[29]. M Peura, J Iivarinen,: Efficiency of simple shape descriptors, Proceedings of Third International Workshop on Visual form (IWVF3), 1997.

[30]. C Cheng, W Liu, H Zhang,: Image Retrieval based on Region Shape, Proceedings of 13th SPIE symposium on Electronic imaging, Storage and Retrieval for image and Video Database, 2001.

[31]. A. Soffer, H. Samet,: Negative Shape Features for Image Database Consisting of Geographic Symbols, Third International Workshop on Visual Form Capri, Italy, 1997.

[32]. D Zhang, and G lu,: A Comparative Study of Fourier Descriptors for Shape Representation and Retrieval, 5th Asian Conference on Computer Vision,(ACCV) , 2002.

[33]. H Kauppinen, T Seppanen, M Pietikainen,: An experimental Comparison of Autoregressive and Fourier-based Descriptors in 2D Shape Classification, IEEE transactions on Patter Analysis and Machine Intelligence, 1995.

[34]. R B Yadava, N K Nishchala, A K Gupta, and V K Rastogi,: Retrieval and Classification of Shape based Objects using Fourier , Generic Fourier, and Wavelet Fourier Descriptors Technique: A Comparative Study, Optics and Laser Engineering, Elsevier, 2007.

[35]. D Zhang and G Lu, A Comparative Study on Shape Retrieval using Fourier Descriptors with different Shape Signatures, Proceedings, International Conference on Intelligence Multimedia and Distance Education (ICIMADE01), 2001.

[36]. Y P Wang, K T Lee, K Toraichi,: Multiscale curvature-based shape representation using B-spline wavelets, IEEE Transactions on Image Processing,1999.

[37]. F Mokhtarian, and A K Mackworth,: A Theory of Multiscale, Curvature based Shape representationfor planar curves, IEEE Transactions of Pattern Analysis and Machine Intelligence, 1992.

[38]. A C Jalba, M H F Wilkinson, J B T M Roerdink,: Shape representation through morphological curvature scale spaces, IEEE Transactions on Image Processing, 2006.

[39]. M Sonka, Hlavac, R Boyle,: Image Processing, Analysis and Machine Vision, Chapman and Hall, London, Uk, 1993.

[40]. M K Hu,: Visual Pattern Recognition by moments Invariants, IRE Transactions on Information Theory, 1962.

[41]. G Taubin, and D Cooper,: Recognition and Positioning of rigid objects using algebraic moments invariants, In SPIE Conference on Geometric Methods in Computer Vision, 1991.

[42]. M E Celebi, and Y A Aslandogan,: A Comparative Study of Three Moments based shape descriptors, Proceedings of International Conference of Information Technology: Coding and Computing, 2005.

[43]. R Mukundan,: A New class of rotational invariants using discrete orthogonal moments, International Conference on Signal Image processing, 2004.

[44]. E. Momeni, D. J. Armaghani, M. Hajihassani, M. F. M. Amin : Prediction of uniaxial compressive strength of rock samples using hybrid particle swarm optimization-based artificial neural networks. Measurement 2015, 60, 50–63.

[45]. L. Wu, and Y. Zhang : Classification of Fruits Using Computer Vision and a Multiclass Support Vector Machine. Sensors 2012, 12, 12489–12505.

[46]. V. Chandwani, V. Agrawal, R. Nagar : Modeling slump of ready mix concrete using genetic algorithms assisted training of Artificial Neural Networks. Expert Syst. Appl. 2015, 42, 885–893.

[47]. S. M. Awan, M. Aslam, Z. A. Khan, H. Saeed : An efficient model based on artificial bee colony optimization algorithm with Neural Networks for electric load forecasting. Neural Comput. Appl. 2014, 25, 1967–1978.

[48]. A. Rocha, D. C. Hauagge, J. Wainer, and S. Goldenstein : Automatic fruit and vegetable classification from images, Comput. Electron. Agriculture, vol. 70, no. 1, pp. 96–104, Jan. 2010.

[49]. S. R. Dubey and A. S. Jalal : Robust approach for fruit and vegetable classification," Procedia Eng., vol. 38, pp. 3449–3453, 2012.

Ranjit K N received Bachelor's degree from VTU, Belgaum and Master's degree from, University of Mysore, Karnataka, India. Her research interest includes image processing, pattern recognition and video understanding. Currently she is working as Assistant professor at MIT, Thandavapura Mysore Karnataka. She has guided several masters' students for their projects in the area of Computer Cognition Technology. She has published many papers in International Conferences and Journals

Raghunandan K S received masters from University of Mysore in the Year of 2013. Currently, he is pursuing Ph.D. at University of Mysore, Karnataka, India. His research interest includes image processing, pattern recognition and video understanding. He has published many papers in International conferences and Journals.

Naveen C received Bachelor's, Master's and Doctorate degree from VTU, Belgaum, Karnataka, India. His research interest includes image processing, pattern recognition and video understanding. Currently working as Professor IN THE Department of Computer Science & Engineering, SJBIT Bangalore. He has guided four Ph.d Students and several masters' students for their projects. He has published many papers in International conferences and Journals. He has delivered many research talks and presented symposium across the globe.

Chethan H K received Bachelor's, Master's and Doctorate degree from University of Mysore, Karnataka, India. Currently working as Professor at Maharaja Institute of Technology, Thandavapura, Karnataka India. Guiding eight Ph.d Students in several domains. Have guided several projects for bachelors and masters' student. He has published many papers in International conferences and Journals.

Sunil C received degree in Bachelor of Electrical and Electronics Engineering and M.Tech in Bio-medical Signal processing and Instrumentation from Visvesvaraya Technological University, Belgaum, Karnataka. Currently he is pursuing Ph.D. at University of Mysore, Karnataka. His research interest includes image processing, pattern recognition, video understanding and Bio-medical Image Processing.