Species and variety detection of fruits and vegetables from images

Shiv Ram Dubey and Anand Singh Jalal*

GLA University,
17 km Stone, NH-2, Mathura-Delhi Highway,
P.O. Chaumuhan, Mathura-281 406 (U.P.), India
E-mail: shivram1987@gmail.com
E-mail: asjalal@gla.ac.in
*Corresponding author

Abstract: Efficient detection of 'species and variety' of fruits and vegetables from the images is one of the major challenges for the computers. In this paper, we introduce a framework for the fruit and vegetable classification problem which takes the images of fruits and vegetables as input and returns it is species and variety. The input image contains fruit or vegetable of single variety in arbitrary position and in any number. This paper also introduces a texture feature based on sum and difference of intensity values of the neighbouring pixels of the colour images. The experimental results show that the proposed texture feature supports accurate fruit and vegetable recognition and performs better than other state-of-the-art colour and texture features. The classification accuracy for the proposed ISADH texture feature is achieved up to 99%.

Keywords: multiclass SVM; sum and difference histogram; global colour histogram; GCH; colour coherence vector; CCV.

Reference to this paper should be made as follows: Dubey, S.R. and Jalal, A.S. (2013) 'Species and variety detection of fruits and vegetables from images', *Int. J. Applied Pattern Recognition*, Vol. 1, No. 1, pp.108–126.

Biographical notes: Shiv Ram Dubey received his BTech in Computer Science and Engineering in 2010 from the Gurukul Kangri Vishwavidyalaya Haridwar, India and his MTech in Computer Science and Engineering in 2012 from GLA University Mathura, India. His research interests are image processing, computer vision and machine learning.

Anand Singh Jalal received his MTech in Computer Science from Devi Ahilya Vishwavidyalaya, Indore, India. He received his PhD in the area of Computer Vision from Indian Institute of Information Technology (IIIT), Allahabad, India. He has 12 years of teaching and research experience and currently, he is working as an Associate Professor in Department of Computer Engineering and Applications, GLA University, Mathura, India. His research interests include image processing, computer vision and pattern recognition.

1 Introduction

In agricultural science, images are the important source of data and information. To reproduce and report such data, photography was the only method until recently. It is difficult to treat or quantify the photographic data mathematically. Digital image analysis and image processing technology circumvent these problems based on the advances in computer and microelectronics associated with traditional photography. This great tool helps to improve images from microscopic to telescopic range and offers a scope for their analysis.

The aim of this paper is to perform automatic image processing in the field of agriculture. Several applications of image processing technology have been developed for the agricultural operations (Nasir et al., 2012). These applications involve implementation of the camera-based hardware systems or colour scanners for inputting the images. We have attempted to extend image processing and analysis technology to a broad spectrum of problems in agriculture. The computer-based image processing is undergoing rapid evolution in parallel with the computing systems. The dedicated imaging systems available in the market, where user can press a few keys and get the results, are not very versatile and more importantly, they have a high price tag on them. Additionally, it is hard to understand how the results are being produced. We develop an efficient solution for the fruit and vegetable classification problem in this paper.

Pattern recognition and contemporary vision problems such as DNA sequencing, fingerprinting identification, image categorisation, and face recognition often require an arbitrarily large number of properties and classes to consider. Recognition system is a ‘grand challenge’ for the computer vision to achieve near human levels of recognition. The work presented in this paper can be used for the automatic designing of agricultural operations through the remote images. The fruit and vegetable classification can be used in the supermarkets where prices can be determined automatically for the fruits purchased by a customer. Fruits and vegetables classification can also be used in computer vision for the automatic sorting of fruits from a set, consisting of different kinds of fruit.

Recognising different kind of fruit and vegetable is a regular task in the supermarkets, where the cashier must be able to identify not only the species of a particular fruit or vegetable (i.e., banana, apple, pear) but also identify its variety (i.e., Golden Delicious, Jonagold, Fuji), in order to determine it is price. This problem has been solved by using barcodes for packaged products but most of the consumers want to pick their product, which cannot be packaged, so it must be weighted. Issuing codes for each kind of fruit and vegetable is a common solution to this problem; but this approach has some problems such as memorisation is hard, which may be a reason for errors in pricing.

As an aid to the cashier, a small book with pictures and codes is issued in many supermarkets; the problem with this approach is that flipping over the booklet is time-consuming. This research reviews several image descriptors in the literature and presents a system to solve the fruit and vegetable recognition problem by adapting a camera at the supermarket that recognise fruits and vegetables on the basis of colour and texture cues. Formally, the system must output a list of possible type of species and variety for an image of fruits or vegetables of single variety, in random position and in any number. Objects inside a plastic bag can add hue shifts and specular reflections.

Given the variety and impossibility of predicting which types of fruits and vegetables are sold, training must be done on site by someone having little or no technical knowledge. The solution of the problem is that the system must be able to achieve higher level of accuracy by using only few training examples.

Image categorisation, in general, relies on the combination of structural, statistical and spectral approaches. Structural approaches describe the appearance of the object using well-known primitives, for example, patches of important parts of the object.

Statistical approaches represent the objects using local and global descriptors such as mean, variance, and entropy. Finally, spectral approaches use some spectral space representation to describe the objects, for example, Fourier spectrum (Gonzalez and Woods, 2007). The proposed method exploits statistical colour and texture properties to classify fruits and vegetables in a multi-class scenario. In this paper, we develop a framework for the fruits and vegetables classification problem to recognise the species and variety of the fruit or vegetable from the images.

There are a number of challenges that must be addressed to perform automatic recognition of the different kind of fruits or vegetables using images from the camera. Many types of fruit and vegetable are subject to significant variation in texture and colour, depending on how ripe they are. Colour and texture plays an important role in visual perception because these are the fundamental characteristic of natural images. Instead of considering colour and texture feature separately, we also introduce an efficient texture feature derived from the colour images.

The rest of the paper is structured as follows: Section 2 gives a brief overview of previous work in object recognition and image categorisation; Section 3 introduces the solution for the fruits and vegetables classification problem as well as a texture feature based on the intensity values of the neighbouring pixels for the image categorisation problems; Section 4 reports the experimental results; and Finally, Section 5 draws the conclusion and future direction.

2 Related works

Recently, a lot of activity in the area of image categorisation has been done. Previous approaches considered patterns in colour, edge and texture properties (Pass et al., 1997; Stehling et al., 2002; Unser, 1986).

Veggie-vision (Bolle et al., 1996) was the first attempt to the fruit and vegetable recognition problem. The system uses texture, colour and density (thus requiring some extra information from the images). This system does not take some advantage of recent developments, because it was created some time ago. The reported accuracy was around 95% in some scenarios but it used the top four responses to achieve such result.

Rocha et al. (2010) presented a unified approach that can combine many features and classifiers. The authors approached the multi-class classification problem as a set of binary classification problem in such a way that one can assemble together diverse features and classifier approaches custom-tailored to parts of the problem. They have achieved good classification accuracy in some scenario but they used top two responses to achieve them. Their method shows poor results for some type of fruit and vegetable such as Fuji Apple.

In general, the fruit and vegetable recognition problem can be seen as an instance of object's categorisation. In Turk and Pentland (1991) employed principal component analysis (PCA) and obtained the reconstruction error of projecting the whole image to a subspace then returning to the original image space. However, it depends heavily on pose, shape and illumination.

A new image descriptor for the image categorisation, the progressive randomisation (PR) was introduced in the literature by Rocha and Goldenstein (2007) that uses perturbations on the values of the least significant bits (LSB) of images. The introduced methodology captures the changing dynamics of the artefacts inserted between a perturbations process in each of the broad-image classes. The major drawback of using PR is that it uses only LSB of the images which lacks the information contained in the most significant bits (MSB) of the images.

Cutzu et al. (2005) used colour, edge, and texture properties for differentiating photographs of paintings from photographs of real scenes. Using single feature they achieved 70–80% correct discrimination performance, whereas they achieved more than 90% correct discrimination results using multiple features.

Low- and middle-level features are used to distinguish broad classes of images (Lyu and Farid, 2005; Serrano et al., 2002). In addition, an approach to establish image categories automatically using histograms, shape and colours descriptors with an unsupervised learning method was presented by Heidemann (2005). Recently, Agarwal et al. (2004) and Jurie and Triggs (2005) adopted approaches that take categorisation problem as the recognition of specific parts that are characteristic of each object class.

Marszalek and Schmid (2006) have extended the category classification with bag-of-features, which represents an image as an orderless distribution of features. They have given a method to exploit spatial relations between the features by utilising object boundaries when supervised training is in progress. They increase the weight of features that agree on the shape and position of the objects and suppress the weight of features that used in the background. But they achieve lower results than expectation in some cases.

Sets of local features which are invariant to image transformations are used effectively when comparing images. These techniques, generally called bag-of-features, showed good results even though they do not attempt to use spatial constraints among features (Grauman and Darrel, 2005; Sivic et al., 2005).

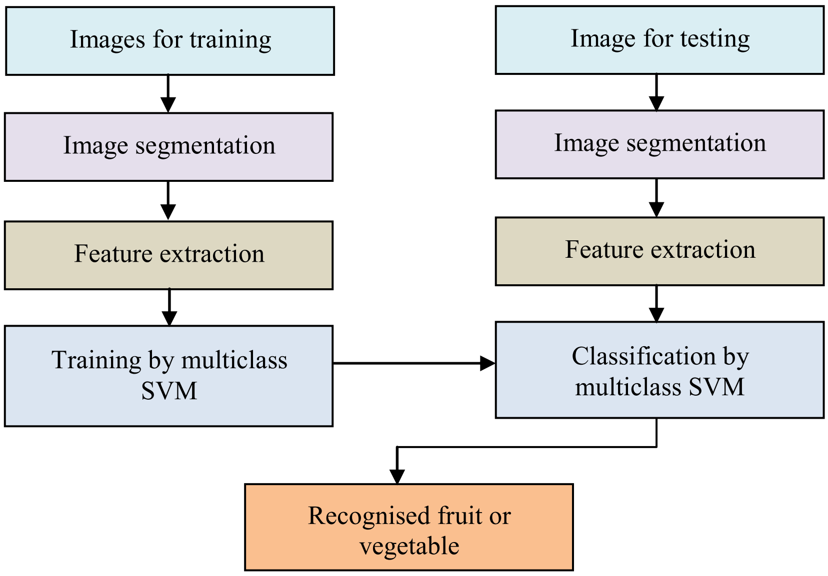
Another interesting technique was proposed by Berg et al. (2005). They have exploited the concept of feature points in a gradient image. By joining a path, the points are connected and a match is finalised if contour found is similar enough to the one presents in the database. A very serious drawback of this method is that it requires a non-linear optimisation step for the finding of best contour; still it relies too heavily on the silhouette cues, which are not very informative cues for fruits like lemons, melons and oranges.

Using a generative constellation model, Weber (2000) has taken spatial constraints into account. The algorithm can work with occlusion in a very good manner, but very costly (exponential with the number of parts). A further work made by Fei-Fei et al. (2006) introduced pre knowledge for the estimation of the distribution, so it reduces the number of examples used for the training to around ten images while having a good recognition rate. The problem of exponential growth in the number of parts makes it impractical for the classification problem presented in this paper.

3 Proposed framework

The proposed framework for the fruit and vegetable recognition system, shown in Figure 1, operates in two phases (i.e., training and classification). Both require some preprocessing (i.e., image segmentation and feature extraction).

Figure 1 Fruit and vegetable recognition system



The presented approach is composed of three steps, in the first step fruit images will be segmented into foreground and background using K-means clustering technique. In the second step features are extracted from the segmented image (i.e., foreground of the image). In the last step fruits and vegetables are classified into one of the classes using trained multi-class support vector machine (MSVM). We also introduce a texture feature to achieve more accurate result for the fruits and vegetables classification in this section.

3.1 *Image segmentation*

Image segmentation is a convenient and effective method for detecting foreground objects in images with stationary background. Background subtraction is a commonly used class of techniques for segmenting objects of interest in a scene. This task has been widely studied in the literature by Rocha et al. (2010). Background subtraction techniques can be seen as a two-object image segmentation and, often, need to cope with illumination variations and sensor capturing artefacts such as blur. Specular reflections, background clutter, shading and shadows in the images are major factors which must be addressed. Therefore, in order to reduce the scene complexity, it might be interesting to perform image segmentation focusing on the object's description only.

For a real application in a supermarket, background subtraction needs to be fast, requiring only fractions of a second to carry out the operation. The best channel to perform the background subtraction is the S channel of HSV-stored images. This is intuitive from the fact that the S channel is much less sensitive to lighting variations than any of the RGB colour channels.

Figure 2 Extracting region of interest from the images, (a) before segmentation (b) after segmentation

(a)

(b)

We use a background subtraction method based on K-means clustering technique (Rocha et al., 2010). Amongst several image segmentation techniques, K-means based image segmentation shows a trade-off between efficient segmentation and cost of segmentation. Some examples of image segmentation are shown in Figure 2.

3.1.1 Algorithm for image segmentation using K-mean

1 I ← Down-sample the image using simple linear interpolation to 25% of its original size.

2 Extract the S channel of I and represent it as 1-d vector V of pixel intensity values.

3 Perform clustering D ← K-means (V, $k = 2$).

4 M ← D back to image space by linear scanning of D.

5 UP ← Up-sample the generated binary M to the input image size.

6 Close small holes on UP using the closing morphological operator with a disk structuring element of radius 7 pixels.

3.2 Feature extraction

In this section, we present an efficient texture feature for the image categorisation problems. Unser (1986) has defined sum and difference histogram of an image which are calculated from the sum and difference of two intensity values with a displacement of ($d1$, $d2$). He considered the displacement in x- and y-directions simultaneously even though there is much information present in the x- and y-directions individually. In this section we improve the Unser's descriptor by considering information present in x- and y-direction separately. To use the information present in both x- and y-directions, first we calculate the sum and difference in x-direction and then simulate this result in the y-direction. Simulation is carried out by taking the sum and difference in y-direction on the outcome of x-direction.

3.2.1 Proposed improved sum and difference histogram texture feature

We introduce an improved sum and difference histogram (ISADH) texture feature which is an improvement of the Unser's descriptor. We also analyse the accuracy of proposed ISADH texture feature and compare with other colour and texture features in the multi-class fruits and vegetables classification scenario in this paper.

3.2.2 ISADH feature algorithm

1. Find the sum S and difference D for the first channel of an image I with a displacement of (1, 0) as:

   $$S(x, y) = I(x, y) + I(x+1, y) \tag{1}$$

   $$D(x, y) = I(x, y) - I(x+1, y) \tag{2}$$

2. Find the sum $S1$ and difference $D1$ of S with a displacement of (0, 1) as:

   $$S1(x, y) = S(x, y) + S(x, y+1) \tag{3}$$

   $$D1(x, y) = S(x, y) - S(x, y+1) \tag{4}$$

3. Find the sum $S2$ and difference $D2$ of D with a displacement of (0, 1) as:

   $$S2(x, y) = D(x, y) + D(x, y+1) \tag{5}$$

   $$D2(x, y) = D(x, y) - D(x, y+1) \tag{6}$$

4. Find the histogram for the first channel by concatenating the histograms of $S1$, $D1$, $S2$, and $D2$.
5. Repeat step1 to step4 for the second and third channel of the colour image.
6. Concatenate the histograms of all three channels in order to find the *ISADH* texture feature of the input image I.

ISADH texture feature relies upon the intensity values of neighbouring pixels. The histogram of two images of the same class may vary significantly. On the other hand, the ISADH feature has less difference for these images. If the difference in feature of two images is less, then images are more likely belongs to the same class. But if the difference is significant, then images are more likely belongs to the different class. This can be illustrated by an example of two 5 × 5 matrix having intensity values in the range of 0 to 7. Let we have two matrices 'A' and 'B' having little difference in their values as follows:

Matrix 'A'

3	5	3	4	6
2	4	2	6	1
2	6	4	1	7
2	4	5	2	6
2	5	4	4	5

Matrix 'B'

3	7	3	4	6
2	3	2	4	1
2	7	4	1	7
2	4	3	2	6
2	3	4	4	3

Calculate three features:

1 simple histogram

2 Unser's feature

3 ISADH feature.

The length of each feature calculated is 8-bin. Table 1 shows the simple histogram for 'A' and 'B', Table 2 shows the Unser's feature for 'A' and 'B', and Table 3 shows the ISADH feature for 'A' and 'B'.

Table 1 Simple histogram for both matrices (8-bin)

Simple histogram	*I0*	*I1*	*I2*	*I3*	*I4*	*I5*	*I6*	*I7*
Matrix 'A'	0	2	6	2	6	4	4	1
Matrix 'B'	0	2	6	6	6	0	2	3

Table 2 Unser feature for both matrices (8-bin)

Unser's feature	*I0*	*I1*	*I2*	*I3*	*I4*	*I5*	*I6*	*I7*
Matrix 'A'	0.5	2.5	7	2.5	0.5	3.5	6	2.5
Matrix 'B'	0.5	5	5	2	0.5	3.5	5.5	3

Table 3 ISADH feature for both matrices (8-bin)

ISADH feature	*I0*	*I1*	*I2*	*I3*	*I4*	*I5*	*I6*	*I7*
Matrix 'A'	0	6.25	2	4.25	0.5	5.75	1.25	5
Matrix 'B'	1.25	5	1.75	4.5	1.25	5	1.5	4.75

In Tables 1, 2, and 3, I0 to I7 represents the intensity levels (i.e., 0 to 7 for 8-bin). Let the difference between the feature of matrix 'A' and the feature of matrix 'B' are defined as the sum of square of difference of the values for each intensity level and can be calculated using the following equation:

$$Diff = \sum_{i=0}^{7} (FA(i) - FB(i))^2 \tag{7}$$

where *FA* is feature of matrix 'A', *FB* is feature of matrix 'B', and *Diff* is the difference between *FA* and *FB*.

Table 4 shows the values of difference for three features Simple Histogram, Unser's feature, and ISADH feature for the 'A' and 'B'. From the Table 4, it is clear that ISADH feature has the lowest value of difference between 'A' and 'B'. The value of *Diff* will be minimal if 'A' and 'B' are more likely belongs to the same class and that is achieved in the case of ISADH feature.

Table 4 Difference in feature of matrix 'A' and matrix 'B'

Feature	*Difference*
Simple histogram	40
Unser's feature	11
ISADH feature	4.5

3.3 Training and classification

Recently, Rocha et al. (2010) presented a unified approach that can combine many features and classifiers. The author approaches the multi-class classification problem as a set of binary classification problem in such a way one can assemble together diverse features and classifier approaches custom-tailored to parts of the problem. A class binarisation is defined as a mapping of a multi-class problem onto two-class problems (divide-and-conquer) and binary classifier is referred as a base learner. For N-class problem $N \times (N - 1) / 2$ binary classifiers will be needed where N is the number of different classes.

The ij^{th} binary classifier uses the patterns of class i as positive and the patterns of class j as negative. The minimum distance of the generated vector (binary outcomes) to the binary pattern (ID) representing each class yields the final outcome. Test case will belong to that class for which the distance between ID of that class and binary outcomes will be minimum.

Table 5 Unique ID of each class

	$x \times y$	$x \times z$	$y \times z$
x	+1	+1	0
y	–1	0	+1
z	0	–1	–1

The approach can be understood by a simple three class problem. Let three classes are x, y, and z. Three binary classifiers consisting of two classes each (i.e., $x \times y$, $x \times z$, and $y \times z$) will be used as base learners, and each binary classifier will be trained with training images. Each class will receive a unique ID as shown in Table 5. To populate the table is straightforward. First, we perform the binary comparison $x \times y$ and tag the class x with the outcome +1, the class y with –1 and set the remaining entries in that column to 0. Thereafter, we repeat the procedure comparing $x \times z$, tag the class x with +1, the class z with –1, and the remaining entries in that column with 0. In the last, we repeat this procedure for binary classifier $y \times z$, and tag the class y with +1, the class z with –1, and set the remaining entries with 0 in that column, where the entry 0 means a 'Do not care' value. Finally, each row represents unique ID of that class (e.g., $y = [-1, +1, 0]$).

Each binary classifier results a binary response for any input example. Let's say if the outcomes for the binary classifier $x \times y$, $x \times z$, and $y \times z$ are +1, –1, and +1 respectively then the input example belongs to that class which has the minimum distance from the vector [+1, –1, +1]. So, the final answer will be given by the minimum distance of:

$$\min \ \text{dist}(\{+1,-1,+1\},(\{+1,+1,0\},\{-1,0,+1\},\{0,-1,-1\}))$$

In this experiment, we have used MCSVM as a set of binary support vector machines (SVMs) for the training and classification.

4 Results and discussion

In this section, we describe the data set of fruits and vegetables, evaluate the proposed approach over the 15 types of fruits and vegetables and discuss various issues regarding the performance and efficiency of the system. In the Section 4.1, we describe the data set used in this experiment and highlight several difficulties present in the data set. In the Section 4.2, the performance of proposed ISADH texture is presented and compared with other colour and texture feature. In order to show the efficiency of the proposed texture feature, we have compared it with four state-of-the-art features. We also consider and compare the performance of the system in two colour-spaces (i.e., RGB and HSV colour-space).

4.1 Data set

To demonstrate the performance of the proposed approach, we have used a supermarket data set of fruits and vegetables, which comprises 15 different categories: Plum (264), Agata Potato (201), Asterix Potato (181), Cashew (210), Onion (75), Orange (103), Taiti Lime (104), Kiwi (157), Fuji Apple (212), Granny-smith Apple (155), Watermelon (192), Honeydew Melon (145), Nectarine (247), Spanish Pear (158), and Diamond Peach (211): totalling 2,615 images. Figure 3 depicts the classes of the data set. (Data set is available at http://www.ic.unicamp.br/~rocha/pub/downloads/tropical-fruits-DB-1024x768.tar.gz).

Figure 3 Data set used

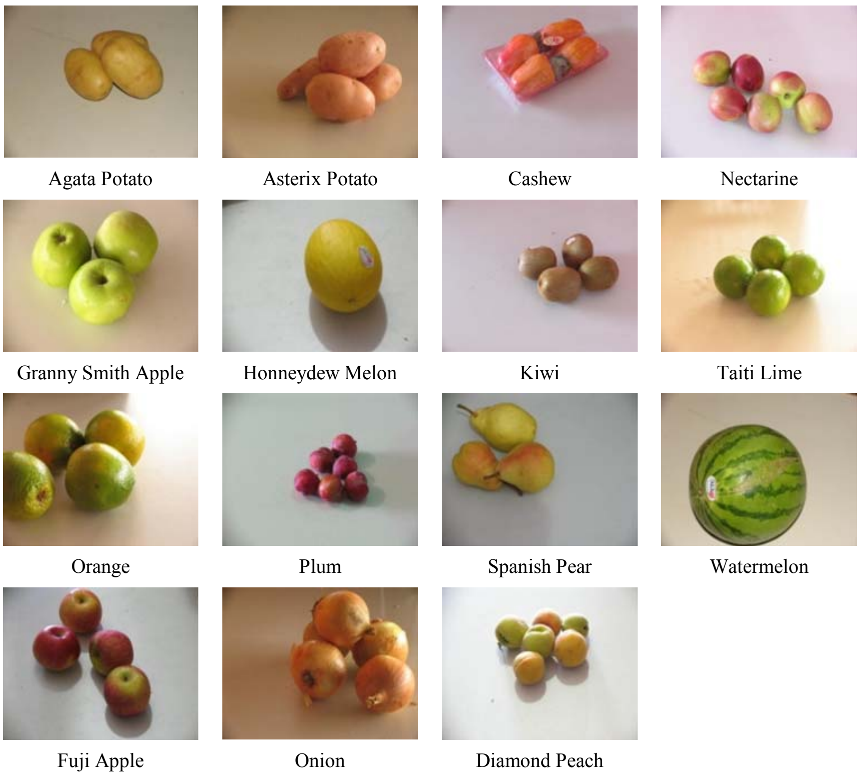


Figure 4 shows an example of kiwi category with different lighting. Figure 5 shows examples of the cashew category with differences in pose. Figure 6 shows the variability in the number of elements for the orange category. Figure 7 shows the examples of cropping and partial occlusion. Presence of these features makes the data set more realistic.

Figure 4 Illumination differences, kiwi category

Figure 5 Pose differences, cashew category

Figure 6 Variability on the no. of elements, orange category

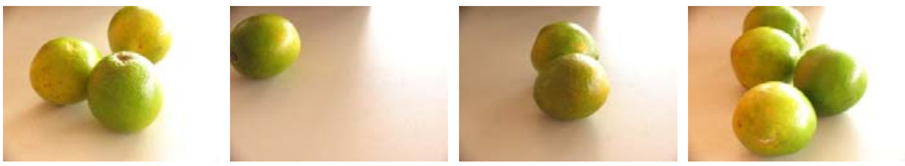


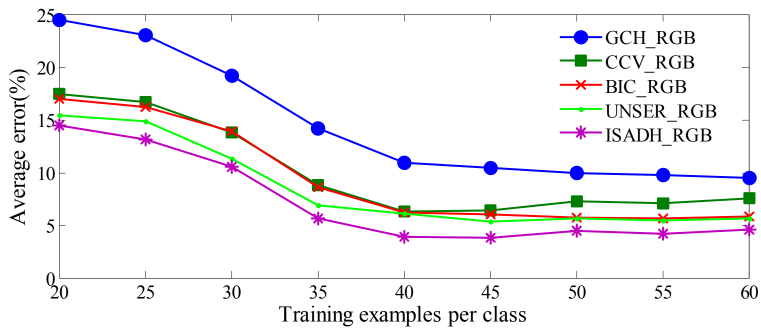
Figure 7 Examples of cropping and partial occlusion

4.2 Result discussion

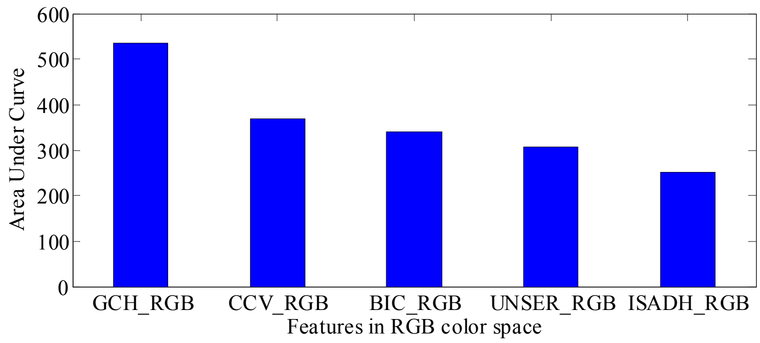
To evaluate the accuracy of the proposed approach quantitatively, we compare our results with the results of global colour histograms (GCH) (Gonzalez and Woods, 2007), colour coherence vectors (CCV) (Pass et al., 1997), border/interior classification (BIC) (Stehling et al., 2002) and Unser's descriptor (Unser, 1986).

GCH is a set of ordered values, one for each distinct colour, representing the probability of a pixel being of that colour. Uniform quantisation and normalisation are used to reduce the number of distinct colours and to avoid scaling bias (Gonzalez and Woods, 2007).

Figure 8 (a) Comparison of GCH, CCV, BIC, Unser, and proposed ISADH features using SVM as a base learner in RGB colour space (b) Area under curve of average error for features in RGB colour space



(a)



(b)

In order to compute the CCV, the method finds the connected components in the image and classify the pixels within a given colour bucket as either coherent or incoherent. After classifying the image pixels, CCV computes two colour histograms: one for coherent pixels and another for incoherent pixels. The two histograms are stored as a single histogram.

In order to find the BIC, the method classifies image pixels as border or interior. A pixel is classified as interior if its 4-neighbours (top, bottom, left, and right) have the same quantised colour. Otherwise, it is classified as border. After the image pixels are classified, two colour histograms are computed: one for border pixels and another for interior pixels. The two histograms are stored as a single histogram.

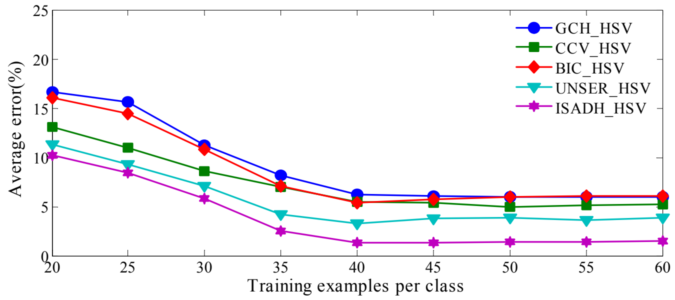
In order to extract the Unser feature, first the method finds the sum and difference of intensity values over a displacement of ($d1$, $d2$) of an image, then it calculates two histograms sum and difference histograms and stores both the histograms as a single histogram.

In this experiment, we use different number of images per class for the training. Figures 8 and 9 shows the average error detected while testing the introduced system for the fruits and vegetables classification considering different features. The x-axis represents the number of images per class for the training and y-axis represents the average error. The average error is computed by the following equation:

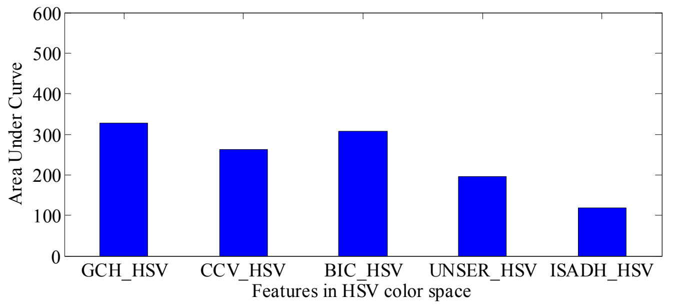
$$Average\ error\ (\%) = \frac{Total\ number\ of\ misclassified\ images}{Total\ number\ of\ images\ used\ for\ testing} \times 100 \tag{8}$$

We have calculated the average error for each feature in RGB and HSV colour space. The average error for each feature in RGB colour space is shown in Figure 8(a) and its area under curve (AUC) is plotted in Figure 8(b).

Figure 9 (a) Comparison of GCH, CCV, BIC, Unser, and proposed ISADH features using SVM as a base learner in HSV colour space (b) Area under the curve of average error for features in HSV colour space



(a)



(b)

The result illustrates that GCH has the highest average error because, it has only the colour information and it does not consider the relation among the neighbouring pixels. The average error for the CCV is less than the average error for the GCH feature because CCV feature exploits the concept of coherent and incoherent regions. BIC feature has low average classification error than the CCV feature because BIC feature takes the values of 4-neighbouring pixel into account.

Unser feature has the low average classification error than GCH, CCV, and BIC. Figures 8 and 9 illustrate that proposed ISADH outperform the other features because ISADH feature considers not only the sum and difference of neighbouring pixel but also the neighbouring information of sum and difference of neighbouring pixel.

Figure 9(a) depicts the average classification error for the features derived from HSV colour images and AUC for Figure 9(a) is plotted in Figure 9(b). GCH feature achieves the highest classification error in HSV colour space also. Both BIC and CCV features have low average classification error than GCH feature. Unser feature performs better in HSV colour space also than GCH, CCV, and BIC features. ISADH feature also outperform and shows highest average accuracy among all features.

We have also evaluated our proposed ISADH texture feature using k-nearest neighbour (KNN) classifier, as shown in Figure 10. The classification accuracy of ISADH feature is better for SVM classifier as compared to KNN classifier. The classification accuracy is computed by the following equation:

$$\textit{Classification Accuracy}\ (\%) = \frac{\textit{Total number of correctly classified images}}{\textit{Total number of images used for testing}} \times 100 \quad (9)$$

Figure 10 Performance of proposed ISADH texture feature in RGB and HSV colour space considering SVM and KNN classifier

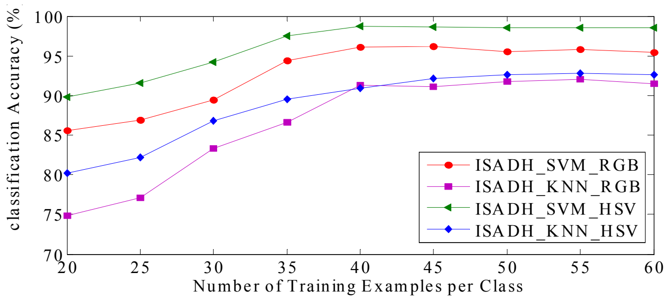
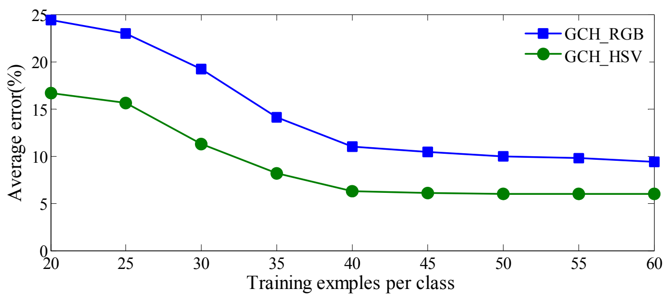
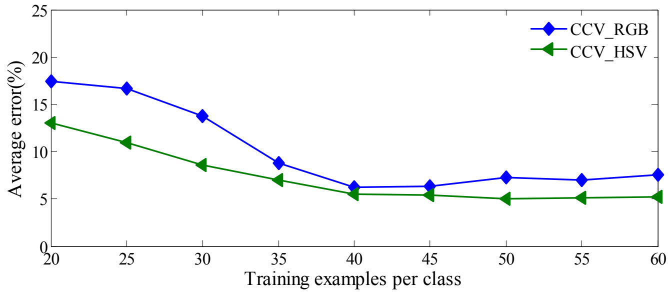


Figure 11 Comparison of results in RGB and HSV colour space using, (a) GCH (b) CCV (c) BIC (d) Unser (e) ISADH feature

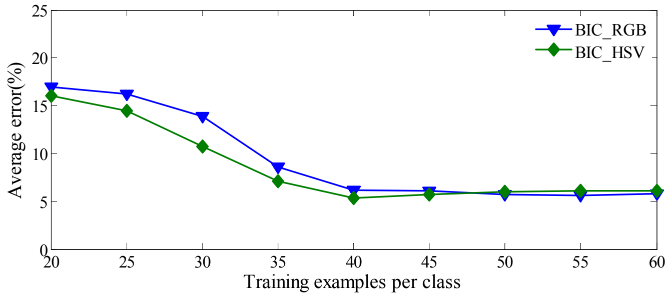


(a)

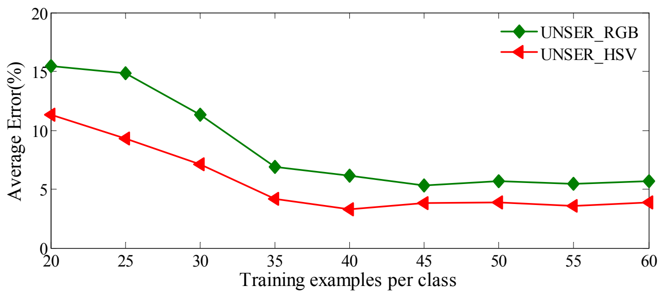
Figure 11 Comparison of results in RGB and HSV colour space using, (a) GCH (b) CCV (c) BIC (d) Unser (e) ISADH feature (continued)



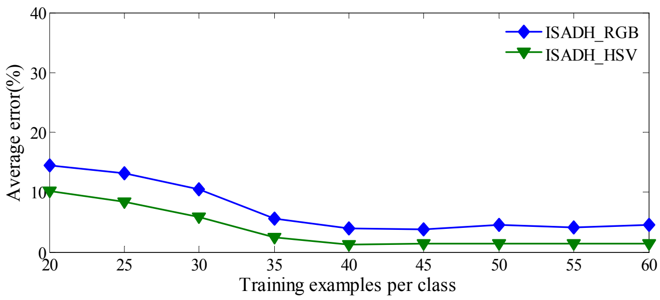
(b)



(c)



(d)



(e)

We have also observed across the plots that HSV colour space is better than the RGB colour space. One possible explanation is that S channel of HSV colour space is very less sensitive to the illumination differences. As shown in Figure 11, all the features perform better in HSV colour space. Table 6 shows the accuracy (%) of fruits and vegetables classification problem when features are extracted from HSV colour images and training is done with 40 training images per class. From this table it is clear that ISADH feature shows a better result in the case of Fuji Apple whereas other features fail to produce good result in the case of Fuji Apple.

Table 6 Accuracy (%) in HSV colour space when system is trained with 40 images per class

S. no.	*Fruit and vegetable*	*No. of test images*	*Accuracy (%) in HSV colour space when system is trained with 40 images per class*				
			GCH	*CCV*	*BIC*	*UNSER*	*ISADH*
1	Agata Potato	161	100	100	99.38	100	100
2	Asterix Potato	141	100	100	100	100	100
3	Cashew	170	100	100	100	100	100
4	Diamond Peach	171	98.25	98.83	98.83	97.66	100
5	Fuji Apple	172	45.93	45.93	47.67	70.35	95.93
6	Granny Smith Apple	115	100	100	100	100	100
7	Honneydew Melon	105	98.10	99.05	99.05	100	98.10
8	Kiwi	117	97.44	98.29	96.58	98.29	98.29
9	Nectarine	207	91.30	94.69	97.58	97.10	97.10
10	Onion	35	94.29	100	100	100	100
11	Orange	63	90.48	96.83	100	95.24	98.41
12	Plum	224	97.77	98.66	99.11	99.11	99.11
13	Spanish Pear	118	96.61	98.31	99.15	96.61	96.61
14	Taiti Lime	64	100	98.44	100	100	100
15	Watermelon	152	98.68	100	100	100	100
	Average accuracy (%)		*93.62*	*95.27*	*95.82*	*96.96*	*98.90*

Figure 12 Some difficult images that are correctly classified

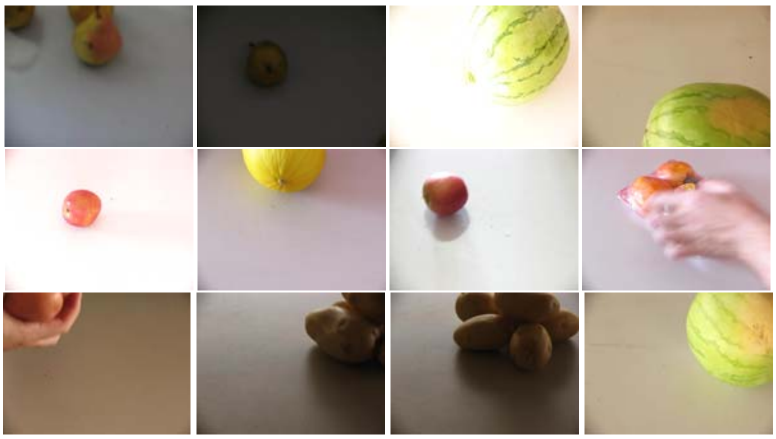


Figure 13 Seventeen images of the data set which are misclassified when SVM is trained with 40 images per class

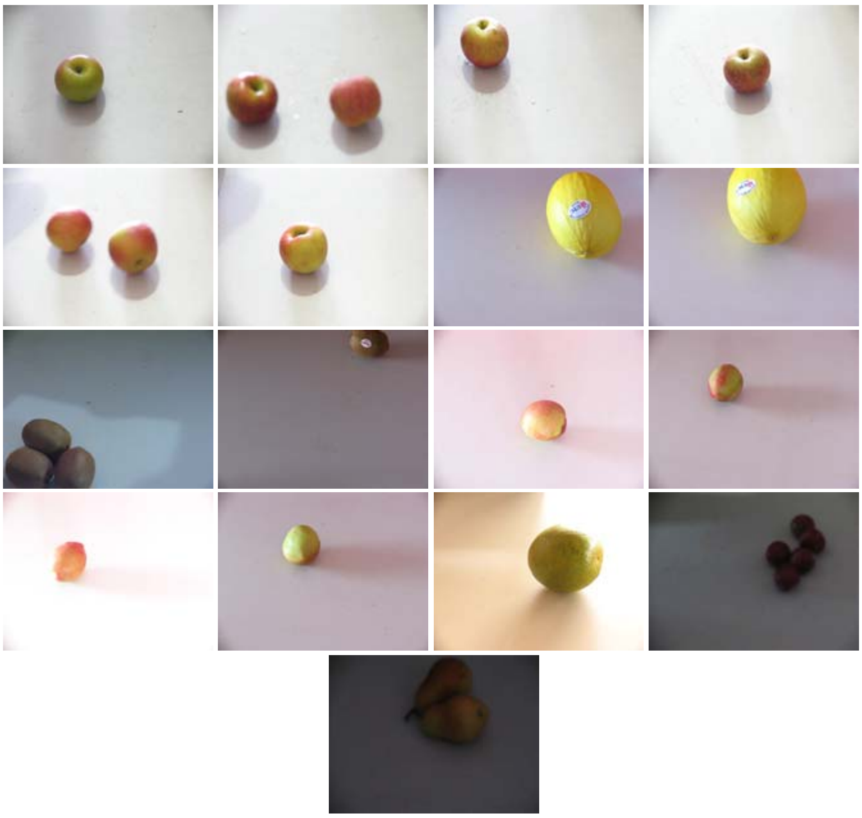


Figure 12 depicts some examples of the fruit and vegetable images which are very difficult to recognise but our method is able to correctly detect the type of fruit or vegetable present in the image. In two images of Figure 12, there is presence of hand within the images but K-means based image segmentation method is able to segment out the portion of hand present in the image. Rest of the images of Figure 12 is either subjected to blurring effect, or cropping effect or illumination differences whereas it is correctly classified by our proposed method.

There are total seventeen images in our data set which are misclassified by our approach when the training is done with 40 images per class as shown in Figure 13. Most of the misclassified images are either of Fuji Apple or Nectarine because the colour and texture of Fuji Apple and Nectarine are very much similar and it is difficult to categorise between them even by human in some cases. Rest of the images of Figure 13 is subjected to either blurring effect, or cropping effect or illumination variation that leads to a misclassification.

5 Conclusions

A framework for the fruits and vegetables classification is introduced and evaluated in this paper. We also proposed an efficient texture feature from the sum and difference of intensity values of neighbouring pixel. The described framework operates in three phases, image segmentation, feature extraction and training and classification. In the proposed approach, image segmentation is done using K-means clustering technique and we compute the ISADH feature for each channel of the colour image and combine these to make a single histogram. The fusion of neighbourhood information with the colour information makes this feature more discriminative than any other colour and texture feature individually. This paper uses a MCSVM for the training and classification. This paper also compared the performance of ISADH for support vector machine and nearest neighbour classifier and indicates that support vector machine is better choice training and classification. The experimental results suggest that the introduced method is able to support the accurate classification of fruits and vegetables from the images. The proposed feature is validated for the fruit and vegetable recognition problem and shows fairly more accurate results compared to other features. One of the future directions of this work is to classify the diseases present in the fruits from the images. Fusion of multiple features and consideration of shape features may enhance the output of the recognition system.

References

Agarwal, S., Awan, A. and Roth, D. (2004) 'Learning to detect objects in images via a sparse, part-based representation', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 11, pp.1475–1490.

Berg, A., Berg, T. and Malik, J. (2005) 'Shape matching and object recognition using low distortion correspondences', *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp.26–33, Washington, DC, USA.

Bolle, R.M., Connell, J.H., Haas, N., Mohan, R. and Taubin, G. (1996) 'Veggie vision: a produce recognition system', *Proceedings of Third IEEE Workshop on Applications of Computer Vision*, pp.1–8, Sarasota, USA.

Cutzu, F., Hammoud, R. and Leykin, A. (2005) 'Distinguishing paintings from photographs', *Computer Vision and Image Understanding*, Vol. 100, No. 3, pp.249–273.

Fei-Fei, L., Fergus, R. and Perona, P. (2006) 'One-shot learning of object categories', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 4, pp.594–611.

Gonzalez, R. and Woods, R. (2007) *Digital Image Processing*, 3rd ed., Prentice-Hall, India.

Grauman, K. and Darrel, T. (2005) 'Efficient image matching with distributions of local invariant features', *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp.627–634, Washington, DC, USA.

Heidemann, G. (2005) 'Unsupervised image categorization', *Image and Vision Computing*, Vol. 23, No. 10, pp.861–876.

Jurie, F. and Triggs, B. (2005) 'Creating efficient codebooks for visual recognition', *Proceedings of the Tenth IEEE International Conference on Computer Vision*, Vol. 1, pp.604–610, Washington, DC, USA.

Lyu, S. and Farid, H. (2005) 'How realistic is photorealistic', *IEEE Transactions on Signal Processing*, Vol. 53, No. 2, pp.845–850.

Marszalek, M. and Schmid, C. (2006) 'Spatial weighting for bag-of-features', *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp.2118–2125, Washington, DC, USA.

Nasir, A.F.A., Rahman, M.N.A. and Mamat A.R. (2012) 'A study of image processing in agriculture application under high computing environment', *International Journal of Computer Science and Telecommunications*, Vol. 3, No. 8, pp.16–24.

Pass, G., Zabih, R. and Miller, J. (1997) 'Comparing images using color coherence vectors', *Proceedings of the Fourth ACM International Conference on Multimedia*, pp.65–73, New York, USA.

Rocha, A. and Goldenstein, S. (2007) 'PR: more than meets the eye', *Proceedings of the Eleventh IEEE International Conference on Computer Vision*, pp.1–8.

Rocha, A., Hauagge, C., Wainer, J. and Siome, D. (2010) 'Automatic fruit and vegetable classification from images', *Computers and Electronics in Agriculture*, Vol. 70, No. 1, pp.96–104.

Serrano, N., Savakis, A. and Luo, A. (2002) 'A computationally efficient approach to indoor/outdoor scene classification', *Proceedings of the 16th International Conference on Pattern Recognition*, Vol. 4, pp.146–149.

Sivic, J., Russell, B., Efros, A., Zisserman, A. and Freeman, W. (2005) 'Discovering objects and their location in images', *Proceedings of the Tenth IEEE International Conference on Computer Vision*, pp.370–377.

Stehling, R., Nascimento, M. and Falcao, A. (2002) 'A compact and efficient image retrieval approach based on border/interior pixel classification', *Proceedings of the Eleventh International Conference on Information and Knowledge Management*, pp.102–109, New York, USA.

Turk, M. and Pentland, A. (1991) 'Eigen faces for recognition', *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, pp.71–86.

Unser, M. (1986) 'Sum and difference histograms for texture classification', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 8, No. 1, pp.118–125.

Weber, M. (2000) *Unsupervised Learning of Models for Object Recognition*, PhD thesis, Caltech, Pasadena, USA.