A New Method for Fruits Recognition System

Woo Chaw Seng
Faculty of Computer Science & Information Technology
University of Malaya

Seyed Hadi Mirisaee
Faculty of Information Science & Technology
University kebangsaan Malaysia

Abstract

Several fruit recognition techniques are developed based upon color and shape attributes. However, different fruit images may have similar or identical color and shape values. Hence, using color features and shape features analysis methods are still not robust and effective enough to identify and distinguish fruits images. A new Fruit recognition system has been proposed, which combines three features analysis methods: color-based, shape-based and size-based in order to increase accuracy of recognition. Proposed method classifies and recognizes fruit images based on obtained features values by using nearest neighbors classification. Consequently, system shows the fruit name and a short description to user. Proposed fruit recognition system analyzes, classifies and identifies fruits successfully up to 90% accuracy. This system also serve as a useful tool in a variety fields such as educational, image retrieval and plantation science.

1. Introduction

Recognition system is an important field of computer science concerned with recognizing patterns, particularly visual and sound patterns[1]. It is central to optical character recognition, voice recognition, and handwriting recognition. It uses methods from statistics, machine learning and other areas. Typical applications are text classification to recognize different type of texts such as spam and non spam E-mails, speech recantation for specified purposes such as translating different languages to English[2], hand written recognition for postal codes, or the automatic face recognition which deals with digital images as input to the pattern recognition systems[3].

In previous years, several types of image analysis techniques are applied to analyze the agricultural images such as fruits and vegetables, for recognition and classification purposes.

The fruits recognition system could be applied as an image contents descriptor which is able to describe the low level visual features or contents of the fruit images for the CBIR system[4]. The most popular analysis techniques that have been used for both recognition and classifications of two dimensional (2D) fruit images are color-based and shape-based analysis methods. However, different fruit images may have similar or identical color and shape values. Hence, using color or shape features analysis methods are still not robust and effective enough to identify and distinguish fruits images.

Therefore, a recognition approach for 2D fruit images is proposed, which combines color-based, shape-based, and size-based methods in order to increase the accuracy of the recognition result. System recognizes provided 2D query fruit image by extracting features values, including color, shape and size and computing extracted features values to measure the distance between the computed features values of query image with the stored standard features values of every fruit example. Fruit Recognition System is an attractive and valuable system that has been developed based on various motivations. Hence, proposed system is developed to research on pattern recognition system, especially on fruits spherical pattern recognition and classification system. In this system, a pattern recognition system is designed that is combination of three different features together, including color, shape, and size to perform sequential pattern classification.

This method can be applied as a useful tool for other object classification and recognition problems. The software solution is able to serve as a useful tool in a variety of fields, such as education, image retrieval, and plant science research. It can be applied for educational purpose to enhanced learning, especially for small kids and Down syndrome patients, of fruits pattern recognition and fruits features classification based on the fruit recognition result. It can be used as a fruit recognition system in grocery store to automate labeling and computing the price. The fruits recognition system could be useful for the plant scientists. The shape and size values of the fruit images that have been computed could assist the plant scientist to do further analysis on variation in morphology of fruit shape in order to help them understand the genetic and molecular mechanisms of the fruits.

Rest of the paper is organized as follows: Section 2 describes some basic definition of fruit recognition. Next section illustrates related works in fruit recognition. Section 4 comprehensively demonstrates proposed method. Section 5 discuss about result and discussion. Conclusion of this paper is given in last section.

2. Background

Computer vision is considered study and application of methods which allows computers to examine and extract image contents or content of multidimensional data in general to facilitate solving a specific vision problem, such as pattern classification problem[5]. There are six main areas of computer vision[5] which are sensing, preprocessing, segmentation, description, recognition, and interpretation. Examples of computer vision applications include systems for controlling processes such as an industrial robot or an autonomous vehicle. Another example is detecting events for visual surveillance or organizing information for indexing databases of images and image sequences. Moreover, modeling objects or environments such as medical image analysis or topographical modeling is another use of computer vision[6]. Computer vision system is also applied for agriculture issues such as a system for monitoring crops growth and weeds under rain shelter[7] or a computer vision system for detecting weeds in cereal crops[8].

Feature extraction is the process of acquiring higher-level information of meaningful object in an image[5]. Analysis with a large number of variables generally requires a large amount of memory and computation power or a classification algorithm, which over fits the training samples and generalizes poorly to new samples. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy [9]. In pattern recognition, features often contain information relative to gray scale, texture, shape or context. In image processing or machine vision, an initial measurement of pattern or some subsequence of measurement pattern is transformed to a new pattern feature[10].

Pattern classification is the process where using for the higher-level information of the object. Thus extracted features are used for assigning the object to a category or a class. It is also used to identify objects in the image automatically by developing the classification algorithm[5]. Pattern recognition is an important field of computer science concerned with recognizing patterns, particularly visual and sound patterns. It uses methods from statistics, machine learning and other areas[11].

The k-Nearest Neighbors is an algorithm that being used widely for classification. During classification process, the unknown object in the query image will be compare to every example of objects that were previously being used to train or develop the classification algorithm[5]. Euclidean distance measures distances in algorithm. [12] The Euclidean distance between two point

$P = (p_1, p_2, p_3, \ldots, p_n)$

and $Q = (q_1, q_2, q_3, \ldots, q_n)$

In Euclidean n-space, is defined as:

$$(1) \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \ldots + (p_n - q_n)^2}$$

3. Related works

Three relevant existing systems that analyze and recognize fruits using shape-based and color-based analysis method have been reviewed. The existing systems, fruits shape variation analyzer, fruits recognition and "localization system and tree fruit recognition using texture properties and color data", are presented in the following sections.

Fruits Shape Variation Analyzer for tomato and other plant species has been developed to analyze shape and size of tomato fruit and other plant species[13], such as Butternut squash, yellow squash, large jalapeno, banana pepper, chili pepper, grape, strawberry, and Bartlett pear. It can accurately determine boundaries of fruit with different color. The Analyzer can perform detection for tomato fruit shape variation and describe any other two-dimensional fruit shape. It provides intuitive descriptors and output that facilitates the analysis of fruit morphology.

The controlled vocabulary and mathematical descriptors were combined into tomato analyzer to achieve objective measurement of fruits shape trait. The controlled vocabulary can accurately describe a range species of fruit shape traits. The mathematical descriptors can calculate the fruit traits with a single equation for each trait, include fruit shape, comprises fruit shape index, fruit shape triangle, fruit shape eccentric, fruit end shape, fruit shape heart, circular, ellipsoid, rectangular; and fruit size comprises fruit height, width, mass, area, and perimeter.

The analyzer contains function for manually adjustment the distal and proximal ends of the fruit. Therefore, no matter fruit are positioned at an angle, the analyzer also can correctly identify the distal and proximal ends of the fruit. This method accuracy is different based on the fruit image but the overall is around 80% to 85% [14].

An AccuRange4000-IR point laser-range finder and a phase shift laser technique have been applied for the fruits recognition and localization to perform automatic selective harvesting[12]. The system uses both of the range and amplitude images provide by the laser range-finder scanner in each scanned scene, and exploits the shape-based methods analysis strategies for fruits recognition and position detection.

The main stages of this shape-based methods recognition strategy are adaptive image smoothing, primitive generation, parameter and evidence estimation, hypothesis generation & verification.

Adaptive image smoothing is necessary to filter the Gaussian white additive noise in a range images. Next, primitive generation stage is needed for recognition of spherical objects in the range image. Then, parameter and evidence estimation stage is needed to estimate the sphere parameter, included 3-D position, radius and reflectance of fruit, and to estimate the degree of confidence over that estimation. Lastly, hypothesis generation and verification stage are used to reject the hypotheses that do not have sufficient evidence value. This method can recognize green fruits and Correctness of this method is 80%.

On tree fruit recognition using texture properties and color data[15], a vision based algorithm presents to locate apples in a single image. Texture based edge detection has been combined with redness measures, and area thresholding followed by circle fitting, to determine the location of apples in the image plane. It was shown that redness works for red apples as well as green apples. This increased texture contrast helped to identify apples separately from background. The algorithm worked equally well for close ups as well as distant images of apples. Results show that the accuracy of system is around 90%.

4. Methodology

The methodology is the method or type of algorithm that being used to develop a system. This section outlines the methodology and data that are used to develop the Fruit Recognition System, and presents the pseudo code for developed system. For Fruits Recognition System, the KNN algorithm performs fruit classification by using the distance measure, which is the Euclidean distance metric to measure the distance between the attributes of the unknown fruit with the stored fruit examples, then the algorithm will find out the nearest or closest examples to the unknown fruit.

The algorithm will prompt user to crop the fruit area in order to obtain or extract the color value of the input fruit image. In order to obtain or compute the fruit roundness value, the algorithm will analyze and extract the fruit region feature properties, thus the fruit area and perimeter for fruit shape roundness or metric value calculates. Algorithm will compute the new area and perimeter values for the fruit image according to the user selected scalar value.

4.1. K-Nearest Neighbors Algorithm

The k-Nearest Neighbors algorithm is the methodology that has been used to develop the Fruit Recognition System. The Fruit Recognition System using the KNN algorithm as a classifier to classify fruit based on mean color values, shape roundness value, area and perimeter values of the fruit.

In order to obtain the area and perimeter of fruit in an image, the image analysis tasks, like preprocessing and segmentation process should be performed on the fruit image.

The fruit area and perimeter are being chosen to represent the fruit size features, which are needed as one of the features to distinguish one type of fruit to another. The area and perimeter values of fruit are estimated in term of pixels values. Meanwhile, the fruit perimeter can be estimated by counting the boundary pixels that have been detected or identified previously; whereas the area of a fruit can be estimated by counting the total number of pixels that are enclosed by the detected fruit boundary area. In order to obtain or compute the fruit size values the algorithm will prompt the user to select a scalar value to resize the fruit image so that its size is approximately the size of fruit in real time environment. After selecting a scalar value, the algorithm will compute the new area and perimeter values for the fruit image.

After inserting training data, system ready to use. The Euclidean distance determines the distances between the features values of the test input fruit with stored training fruits; whereas KNN algorithm will find out the 'K' shortest distance or closest examples to the input fruit then assign the input fruit to the class where majoring of the 'K' closest examples is from. For Fruits Recognition System, the value chosen for 'K' is '1', which means classification of input fruit sample is based on where is the class of the closest fruit example is from.

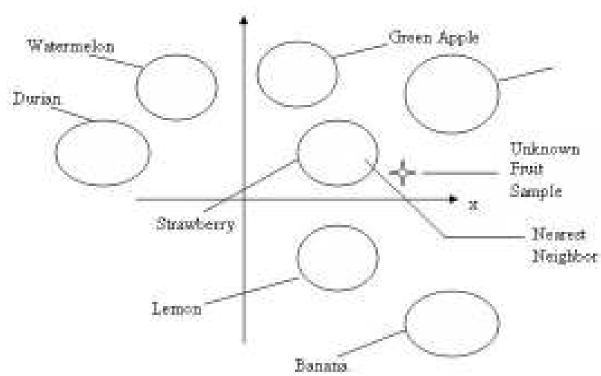

Figure 1: The fruit images space that illustrates the situation and position of the unknown fruit sample and the groups of stored fruit examples during the classification process of unknown fruit sample to the class of its nearest neighbor.

The Fruits Recognition System using the following function [16] to classify the input fruit sample:

Fruit_Class = knnclassify (sample, training, group, k)

This function will classify the input fruit sample by determining distance between the attributes of input fruit sample with attributes of all other training fruit examples and find out the 'k' nearest examples, then classify the unknown input fruit image to the class or group where majoring of the 'k' nearest neighbors is from. Figure 2 illustrates how the situation and position is for the input or unknown fruit sample and for all the fruit examples that have been stored during system training.

4.2. Data

Fifty fruit images have been collected for Fruits Recognition System. These fruit images have been divided into training fruit set and testing fruit set, where 36 of the collected fruit images were used to develop and train the system; whereas 14 of the fruit images were used to test the system. The training fruit images are required to be sent in and processed by the system when develop the classification algorithm for the recognition system.

The mean color values of the fruit can be computed after the user crop the area of fruit in the fruit image. The system will compute the mean values for each of the red, green and blue (RGB) component of the cropped fruit area by manipulating and computing on the 3D matrices that stored all of the fruit pixels. Next the RGB values for each fruit pixel computed by using the mean function provided in MATLAB[16].

The fruit shape roundness or metric values can be computed after extract and estimate the area and perimeter of the fruit by using equation as below[5, 16]:

$$shape = 4\pi\left(\frac{Area}{Perimeter^2}\right) \qquad (2)$$

4.3. Analysis Methods

The fruits recognition system consists of five main processing modules, which are, fruit input selection module, fruit color computing module, fruit shape computing module, fruit size computing module, and fruit classification or recognition module. The first processing module of the system will prompt the user to select a fruit image from the fruit selection menu for further recognition process. The fruit color-computing module is necessary to perform the fruit color feature extraction tasks. Subsequently, the fruit shape-computing module will analyze the fruit after that fruit region feature properties has been extracted. Thus the fruit area and perimeter are used in fruit shape roundness or metric value calculation.

Table 1: Stored details of color (RGB) for each type of fruit in System during training

Fruit Name	Minimum			Maximum		
	R	G	B	R	G	B
Red	128.5	14.0	42.0	219.9	98.0	67.0
Green	99.0	145.0	31.0	146.0	189.0	62.0
Strawbe	158.8	21.38	28.2	233.0	56.1	58.2
Banana	168.0	145.8	53.1	251.0	234..5	95.5
Lemon	192.5	148.3	17.6	252.8	218.5	129.1
Durian	84.8	67.0	22.0	197.8	172.0	88.0
waterme	77.8	119.9	39.7	119.6	166.4	77.7

Table 2: The minimum and maximum area and perimeter values for each type of fruit in Fruit Recognition System

Fruit	Area (pixels)		Perimeter(Pixels)	
	mi	max	min	Max
Red Apple	15	48472	694.6	826.1
Green apple	78	17224	472	945
Strawberry	48	10595	306.4	612.8
Banana	44	18332	622.8	1.7500e+3
Lemon	49	11788	414.7	799.7
Durian	19	83700	1.4255e+3	3.6694e+3
Watermelon	33	81228	1.7038e+3	3.8223e+3

Table 3: The minimum and maximum roundness values for each type of fruit in Fruit Recognition System

Fruit	Roundness	
	min	max
Red Apple	0.82	0.91
Green apple	0.84	0.92
Strawberry	0.60	0.69
Banana	0.22	0.39
Lemon	0.72	0.91
Durian	0.44	0.59
Watermelon	0.66	0.89

The fourth module of the fruit recognition system will perform its tasks to compute the fruit size values, thus, the area and perimeter of the fruit based on user selected scalar value and area and perimeter value obtained from the third module. The classification and recognition module is responsible to classify the input or user selected fruit by using KNN algorithm. This module measure the distance between features values of the selected fruit and the features values of the stored fruit test. Afterward the KNN finds out among the stored fruit example that have shortest distance with the input. Later than, the system will identify and assign the class of input fruit. Eventually, system displays the recognition computed feature results to the user. In the rest of this section, you can see the pseudo code for fruit recognition algorithm.

1. Select fruit image
2. Crop the fruit area
3. Compute mean for RGB components
4. Compute shape by threshold segmentation (remove noises, morphological operations)
5. Compute geometrical properties (Area, perimeter)
6. Use KNN and parameters in 3, 4, 5 to classify the image
7. Output Result

5. Experimental results and discussion

Thirty-six number of the collected fruit images were used to develop and train the system; Table 1, 2 and 3 show details of the color, shape, area and perimeter values for each type of fruit that have been stored during system training. These stored color values, shape roundness values, area and perimeter values are being used as standard feature values for comparison and classification of query or input fruit image to the system. The images are shown in appendix 1. The Fruits Recognition System was tested by using the 14 test fruit images. The test results are accurate for the entire testing fruit set. As well, the system is able to recognize all the test fruit images. Table 3 summarizes the recognition results of Fruits Recognition System on the fruit images that are being sent in as input images during testing the system. The table lists out the test results of the system, including fruit name, computed features values, such as mean RGB values, shape roundness values, area and perimeter values. Tested images are shown in appendix 2.

However, the results are greatly affected by the fruit size scalar values, which are selected by user. The system tester plays an important role on determining the accuracy of the recognition results.

For instance, the tester should crop the area of the fruit region excludes the background region on the image in order to extract the correct color feature values of the fruit so that the system can compute a new area and perimeter values for the resized test fruit image. According to the test results, the recognition of the test fruit images is accurate for those fruit size scalar values that are selected based on the size of fruit in real time environment.

In order to prove the accuracy of the recognition system, more test fruit images for each type of fruit in the system domain should be collected and used to test the system. In addition, training images should be consist of fruits in various colors, shapes and sizes, and capture in different angles and positions, so after that it is conceivable to say the system is robust enough and able to recognize any input of fruit images that are being capture in any conditions.

Table 3: The recognition results of color (RGB) on the test fruit images of Fruits Recognition System

Input Fruit Image number	Selected Scalar Value	Test Result				
		Fruit Name	Mean (Pixels) (R,G,B)	Roundness Values	Area (Pixels)	Perimeter (Pixels)
1	2	Red Apple	(166.42,42.97,49.88)	0.88	21260.00	780.87
2	2	Banana	(173.73,150.07,53.90)	0.30	9166.00	875.38
3	6	Durian	(98.03,87.17,23.19)	0.47	67044.00	3294.51
4	2	Strawberry	(186.52,54.59,27.76)	0.69	10328.00	612.78
5	2	Lemon	(212.84,144.87,4.11)	0.86	11788.00	586.16
6	6	Watermelon	(71.19,86.73,39.68)	0.66	81228.00	3037.29
7	4	Green Apple	(155.85,176.28,31.30)	0.88	15636.00	944.26
8	2	Red Apple	(196.13,101.96,71.82)	0.91	17844.00	701.75
9	4	Banana	(213.77,193.35,91.47)	0.39	11936.00	1245.69
10	6	Durian	(157.67,163.88,89.78)	0.44	78162.00	3669.42
10	8	Watermelon	(157.67,163.88,89.78)	0.44	104216.0	4892.57
11	0	Strawberry	(147.57,42.27,48.33)	0.61	4422.00	301.81
12	2	Lemon	(248.78,211.48,125.05)	0.78	11768.00	614.04
13	6	Watermelon	(74.33,117.40,53.53)	0.89	55350.00	2159.69
14	2	Green Apple	96.74,143.79,22.44)	0.90	16436.00	678.64

6. Conclusion

The proposed method can process, analyze, classify and identify the fruits images, which are selected and sent in to the system based on color, shape and size features of the fruit. The KNN algorithm is the appropriate and effective classification algorithm to be used in the Fruits Recognition System. The recognition system that has been developed is able to recognize all the test fruit images which are being selected by user or system tester from the fruit selection menu on the system. The recognition results of the system are accurate up to 90%.

There are some future works should be implement on the Fruits Recognition System in order to improve and enhance the functionality and flexibility of the recognition system for more widely usage. The system should be improved by extending its functions to process and recognize more variety of different fruit images. Besides that, the texture based analysis technique could be combined with the existing three features analysis technique on the system in order to gain better discerning of different fruit images.

7. Acknowledgment

We would like thanks Mr. Liew Jin Yee For doing this work and implementing the application in a user friendly way.

8. References

[1] Richard O. Duda, P.E.H., David G. Stork, *Pattern classification*. second ed. 2001, New York: Wiley.

[2] Gales, M.J.F.L., X.; Sinha, R.; Woodland, P.C.; Yu, K.; Matsoukas, S.; Ng, T.; Nguyen, K.; Nguyen, L.; Gauvain, J.-L.; Lamel, L.; Messaoudi, A., *Speech Recognition System Combination for Machine Translation*, in *IEEE International Conference on Acoustics, Speech and Signal Processing* 2007, IEEE: Hawaii USA. p. 1277-1280.

[3] Group, W. *Pattern recognition*. 2007 [cited 2008 12/11/2008]; Available from: http://encyclopedia.thefreedictionary.com/Pattern%20recognition.

[4] W.K. Jung, I.C.N., *Relevance Feedback in Content-Based Image Retrieval System by Selective Region Growing in the Feature Space.* Signal Processing: Image Communication, 2003. **18**: p. 13.

[5] E.Umbaugh, S., *Computer Vision and Image Processing: A Practical Approach using CVIP tools*. first ed. 1998: Prentice Hall Professional Technical Reference.

[6] WikiPedia. *Computer Vision*. 2007 [cited 2008 11/11/2008]; Available from: http://en.wikipedia.org/wiki/Computer_vision.

[7] A.R.M. Syafiudin, C.C.T., O.M. Isa, *APPLICATION OF COMPUTER VISION SYSTEM TO MONITOR CROPS GROWTH AND WEEDS UNDER RAIN SHELTER*, in *International Symposium on Greenhouses, Environmental Controls and In-house Mechanization for Crop Production in the Tropics and Sub-Tropics*, I.H.R. R. Kamaruddin, N.R. Abdul Hamid Editor. 2006, ISHS Acta Horticulturae: Malaysia.

[8] Pajares, G.T., A. Burgosartizzu, X.-P. Ribeiro, A. , *Design of a computer vision system for a differential spraying operation in precision agriculture using hebbian learning.* Computer Vision, IET, 2007. **1**(3-4): p. 93-99.

[9] *Feature extraction*. [cited April 2, 2007]; Available from: http://en.wikipedia.org/wiki/Feature_extraction.

[10] *Feature Extraction*. [cited 3rd April 2007]; Available from: http://www.photonics.com/directory/dictionary/lookup.asp?url=lookup&entrynum=1837&letter=f.

[11] *Pattern recognition*. [cited 4th November 2008]; Available from: http://encyclopedia.thefreedictionary.com/Pattern%20recognition.

[12] AA.R. Jimenez, R.C., J.L. Pons. *Shape-based Methods for Fruit Recognition and Localization using a Laser Range-Finder*. in *International Workshop on Bio-Robotics*. 1997.

[13] Marin Talbot Brewer, L.L., Kikuo Fujimura, Nancy Dujmovic, Simon Gray, Esther van der Knaap, *Development of a Controlled Vocabulary and Software Application to Analyze Fruit Shape Variation in Tomato and Other Plant Species.* Plant Physiology, 2006. **141**: p. 15-25.

[14] A. R. Jimenez, R.C., and J. L. Pons,, *A survey of Computer Vision Methods for Locating Fruit on Trees*. ASAE,, 2000. **43**: p. 1911-1920.

[15] Zhao, J.T., J. Katupitiya, J. , *On-tree fruit recognition using texture properties and color data*, in *international conference on Intelligent Robots and Systems*. 2005, IEEE: Edmonton Canada. p. 263-268.

[16] Group, T.M.R. *Image Processing Toolbox User's Guide*. 2008 [cited 14th November 2008]; Available from: http://www.mathworks.com/products/image/description1.html.

Appendix 1: Training images

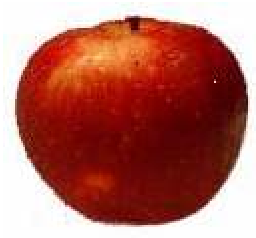

Figure 2: Red Apple

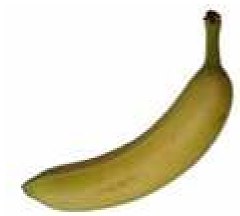

Figure 3: Banana

Figure 4: Lemon

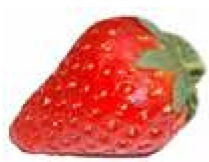

Figure 5: Strawberry

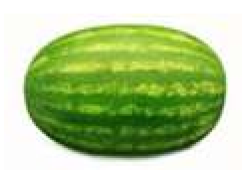

Figure 6: Watermelon

Appendix 2: Tested Images

Figure 7: Red Apple

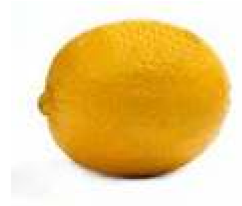

Figure 8: Lemon

Figure 9: Strawberry

Figure 10: Watermelon

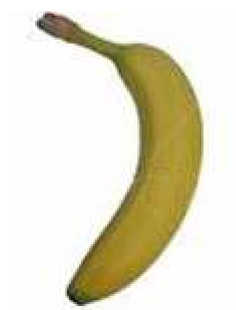

Figure 11: Banana