Transaction of the ASAE, Vol. 43(6): 1911-1920, 2000

A Survey of Computer Vision Methods for Locating Fruit on Trees

A.R. Jiménez, R. Ceres and J.L. Pons
Instituto de Automática Industrial (CSIC)
Ctr. Campo Real, Km. 0.2 La Poveda
28500 Arganda del Rey, Madrid, Spain
Phone: 34 918 711 900 Fax: 34 918 717 050
e-mail: arjimenez@iai.csic.es
http://www.iai.csic.es/lopsi

Abstract

A review of previous studies to automate the location of fruit on trees using computer vision methods was performed. The main features of these approaches are described, paying special attention to the sensors and accessories utilized for capturing tree images, the image processing strategy used to detect the fruit, and the results obtained in terms of the correct/false detection rates and the ability to detect fruit independent of its maturity stage. The majority of these works use CCD cameras to capture the images and use local or shape-based analysis to detect the fruit. Systems using local analysis, like intensity or color pixel classification, allow for rapid detection and were able to detect fruit at specific maturity stages, i.e fruit with a color different from the background. However, systems based on shape analysis were more independent of hue changes, were not limited to detecting fruit with a color different from the color of the background, however their algorithms were more time consuming. The best results obtained indicate that more than 85% of visible fruits are usually detectable, although using CCD sensors there were a certain amount of false detections that in most cases were above > 5%. The approaches using range images and shape analysis were capable of detecting fruit of any color, did not generate false alarms and gave precise information about the fruit three-dimensional position. In spite of these promising results, the problem of total fruit occlusion limits the amount of fruit that can be harvested, ranging from 40 to 100% of total fruit, depending on fruiting and viewing conditions. This fact seriously affects the feasibility of future harvesting robots relying on images that do not contain a high percentage of visible fruit. Therefore, new techniques to reduce total occlusion should be studied in order to make the process feasible.

Keywords: Automatic harvesting, Computer vision, Fruit localization, Color and Shape analysis.

INTRODUCTION

In any production system there is a growing need to obtain higher quality products at a lower cost in order to be competitive. One solution to this challenge is the development of automatic systems that replace manpower in tasks when a person performs worse than an automatic device in terms of precision, repetitivity and working cycle. Probably, harvesting is the process that has received the least amount of technological development for satisfactory automation. Some types of damage resistant agricultural products such as olives and almonds can be harvested using trunk or branch shakers. However the harvesting of delicate fruit, such as oranges, lemons, apples or peaches for the fresh market, is a process that cannot be performed using aggressive methods such as shakers. If these methods were used, the fruits could be damaged by impacting the branches of the tree during the fall, and therefore fruit would not be appropriate for commercialization in the fresh market. The current method for collecting fruit is hand harvesting. This method implies the use of temporal manpower which increases the final cost of the fruit in the market.

The solution for the harvesting of delicate fruit without the drawbacks of manual operation may be the use of automatic systems capable of performing an individualized collection, using selective strategies to harvest only fruit with the desired conditions, and at the same time, providing a system able to work 24 hours a day. To obtain this automatic system there are three main problems to be solved: (1) the guidance of the robot through the crop, (2) the location and characterization of the fruit on the tree, and (3) the grasping and detachment of each piece. The first factor is not crucial and can be bypassed using one operator to guide the robot in the plot. The other two problems have received remarkable attention during the last thirty years, although no commercial harvesting robot is available. This paper presents the main achievements reported to solve the automatic fruit detection problem.

OBJECTIVE

The main objective of this paper is to present the state of the art in the development of automatic systems designed to locate fruits on a tree. To satisfy our goal, the paper deals with the following two topics: (a) the description of the main approaches developed by previous authors to automate the detection and localization of fruits for automatic harvesting purposes; and (b) the analysis and discussion of these results to deduce what main achievements have already been obtained, what remaining problems have not yet been solved which could limit practical applications, and what research efforts are still needed in order to succeed in developing automatic fruit detection systems.

REVIEW OF COMPUTER VISION FOR FRUIT DETECTION

A harvesting robot collecting fruit in an individualized way must be able to guide its mechanical arm towards each piece of fruit on the tree. Therefore, the three-dimensional fruit location must be computed. If a selective harvesting is required, not only its position, but some fruit characteristics like its radius and ripeness have to be estimated. This review is presented in chronological order to show the research evolution in this area.

The first reference considering the automatic detection of fruits dates from 1968 (Schertz and Brown, 1968). In this paper Schertz and Brown suggested that the location of fruits might be determined by photometric information, specifically by using the light reflectivity differences between leaves and fruits in the visible or infrared portion of the electromagnetic spectrum. Schertz and Brown's paper also indicates some problems to be considered in automatic fruit detection: 1) non-uniform illumination (which could be overcome by shading the tree or by using sensors independent of level of illumination), and 2) the foliage concentration that limits the visible fruits to 70-100%. The ideas of Schertz and Brown were applied and refined for inspection purposes by Gaffney (Gaffney, 1969). It was determined that "Valencia" oranges could be sorted by color using a single wavelength band of reflected light at 660 nm. This technique was capable of distinguishing fruits with colors between orange, light orange and green.

The first implemented computer vision system for detecting apples was reported by Parrish and Goksel (1977). This system consisted of a B/W camera and a red optical filter to increase the contrast between red apples and green-colored leaves. Pixel intensity values and connectivity among them were used to perform an image analysis method which was split into three basic steps: (1) an intensity thresholding step to select only the brighter pixels which should belong to the red surfaces; (2) the binary image generated in the previous step was smoothed using morphological filters to eliminate noise and irrelevant details, and finally, (3) for each segment, or region formed by a set of connected pixels, the difference between the lengths of the horizontal and vertical extrema were computed. Accordinly, a roundness measure is obtained as well as the centroid and radius values. The density of the region was then calculated by placing a circular window, whose size was determined by the extrema mean value, over the segment centroid. If the region density was found to be greater than a preset threshold, the region was accepted as an apple. Some experiments were conducted using an artificial apple tree setup but no information about overall detection rates were reported.

Grand D'Esnon (1987) developed a vision system for the MAGALI robot to detect apples using a color camera. An analog signal processing system was able to select points of a given color within the image. Further binary image processing allowed the detection of each segment center. However, this proposed vision method was not robust enough and required the use of a protective covering behind the tree to darken the background, and in this way, to avoid some of the false detection caused by the sky visible through the tree leaves. There was a second version of this system presented by Rabatel (1988) which used three color cameras and the assistance of three different optical filters (950, 650 and 550 nm) to obtain three complementary intensity images. The vision system was based on the analysis of three spectral bands chosen after a spectra-photometry study in the visible and near infra-red bands. This study revealed that there was a narrow spectral band centered at 950 nm where the apple tree's leaves and several apples varieties (Golden Delicious, Red Delicious and Granny Smith which have yellow-green, red and green color, respectively) had similar reflectivity. This feature was used to make the system quite insensitive to illumination changes by referencing the images obtained at 650 and 550 nm to the 950 nm image. The two ratios computed were employed to decide which pixels belonged to a fruit and which could be classified as leaf points. With this strategy it was possible to recognize even green mature apples. The extension of this work to other types of apples or trees involved individual spectral studies for each particular recognition problem, and this fact indicated that the system was very sensitive to changing harvesting conditions. The successful detection rate was approximately 50% and no quantitative data was presented regarding false detections. Nevertheless, the authors declared that some failures were detected and that the system was not absolutely insensitive to illumination changes.

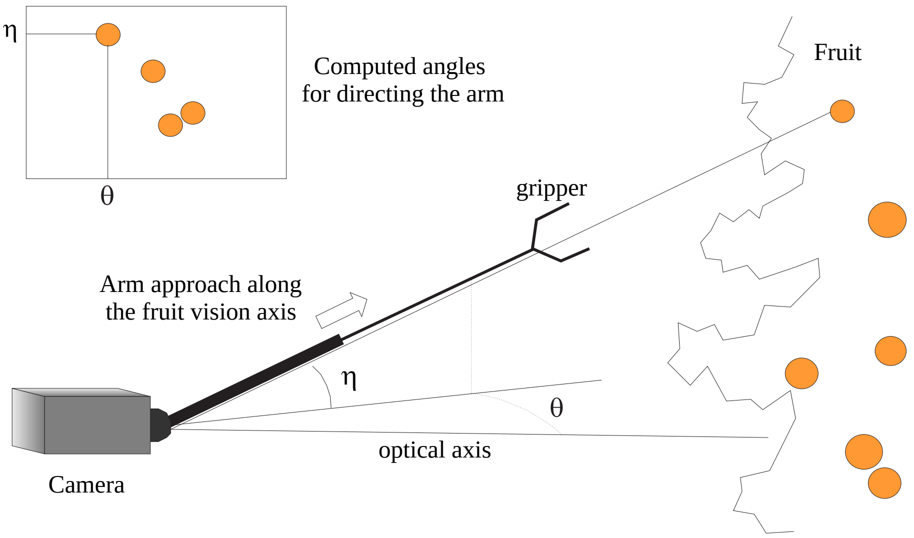

Figure 1: Usual principle to find depth data for a three-dimensional fruit location.

Whittaker (1987) developed a system to recognize and locate tomatoes in natural settings which was insensitive to fruit maturity states, and therefore, insensitive to changes in color. It was considered that analysis methods based on local pixel values were not appropriate to build a processing system insensitive to color changes, so it was proposed to base the analysis on shape information considering not only local data but global relationships within sets of pixels. A B/W CCD camera was used to obtain intensity images with 256 gray levels. Each image was processed using a Sobel kernel, obtaining a map of gradient vectors which was used to obtain an edge image by thresholding and a directional image. This directional image contained vectors indicating, for each edge, the orientation of maximum intensity changes. Using both the edge and directional images, an optimized Circular Hough Transform-CHT (Duda, 1972 and Illingworth, 1988) was applied to detect circular arcs that should correspond to tomato contours. The results obtained were very sensitive to the user-specified threshold value, and the best results for a 99% threshold value were 68% correct detection and 42% false detection. The contour of the leaves was one of the major problems, since the analysis algorithm interpreted them as possible fruits. The authors recognized that, using an 8086 processor or equivalent, the algorithm was computationally intensive and could not be performed in real-time. Obviously, this real-time restriction would not exist nowadays using some of the current computers. Like in previous works, the fruit location was not complete since the distance information was not known and the robotic arm had to be moved toward the fruit using its vision axis until the fruit was contacted or sensed (Fig. 1).

The Italian AID (Agricultural Industrial Development-SpA, Catania) robot vision system was implemented to recognize oranges (Leci, 1998). A color camera with artificial lighting was used. An analog electronic filter enhanced the image and during digitization six bits were used to codify the pixel value which was proportional to the closeness of the actual pixel hue to a preset hue reference. The selected hue reference was close to the orange color searched for, so after digitalizing a pseudo-gray image was obtained where the regions with a color similar to the reference were emphasized. With this pseudo-gray image, a gradient image and a directional image were computed using the Sobel operator. Finally, the scene interpretation was conducted by searching for a match with an object model previously stored. This gradient direction template was moved step by step throughout the directional image. Approximately 70% of the visually recognizable fruit was detected using a 3200 K lamp, but the system performed worse when fruit had a green color. This research, together with the previously described work of Whittaker (1987), was one of the first studies that attempted to recognize spherical forms in the image, in this case through the orientation of gradients.

Slaughter and Harrell, involved in the development of the Citrus Picking Robot (CPR), presented two approaches to solve the fruit recognition problem for the CPR robot and both were based on color information. In the first approach (Slaughter, 1987), a color camera was used without the support of any optical filter, but aided by an artificial lighting system. The Hue and Saturation components of each pixel were used as features to segment the image by applying a traditional classification in a bi-dimensional feature space. The classification was performed with a linear classifier which isolated a rectangular area in the feature space using a maximum and minimum threshold for each feature. Depending on the threshold selected, between 45% and 93% of the pixels were correctly classified. The algorithm running on a Motorola 68020 and analyzing images of 384×485 pixels required 2.5 seconds. Therefore, a hardware implementation using comparators to do the thresholding to increase performance was suggested. In the second approach, Slaughter and Harrell (1989) extended their earlier study by using the RGB components recorded by a color camera without the support of any illumination. Using the RGB components as features and a traditional Bayesian classifier, the image was segmented separating pixels belonging to fruit from those corresponding to the background. Tests demonstrated that 75% of the pixels were correctly classified, obtaining a 100% of correct detection over the total visible fruit. Both approaches, HS and RGB, used color information to perform the classification, so these methods are valid for mature oranges or other fruit easily distinguishable by color from the background.

A very interesting article was published by Harrell (1987), where the different factors affecting the costs of a citrus harvesting robot were analyzed. It was found that robotic harvest costs were mainly affected by harvest inefficiency (15% of fruit not visible), purchase price, pick cycle time, and repair expenses. Further development in the CPR robot was presented by Harrell et al. (1989) to estimate the radius and position of each fruit in the image. This method initially performed a search to detect the segments, and then, an iterative line tracing process to compute the horizontal and vertical diameters. A real-time implementation of the CPR robot was described (Harrel, 1990) where it was concluded that harvest inefficiency is the most limiting factor for a practical solution for robotic harvesting. In fact, it was estimated that robotic harvesting would be 50% more expensive than conventional hand harvesting. Tests conducted for 150 hours in Florida and Italian groves with this harvester prototype allowed them to conclude than there were no false detection problems, because all non-fruit objects in the grove had a different color than citrus colors, and 75% of the fruit detected to be picked was successfully picked. The pick cycle time achieved ranged from 3 to 7 seconds.

Sites (1988) presented a system to recognize ripe apples and peaches. This intensity-based method used a B/W camera and color filters (630 to 670 nm) to increase the contrast between the fruit and the background. Artificial light was used and most of the images were recorded during night operation. The whole method could be divided into five step: (1) thresholding based on an intensity histogram distribution assigning a constant 37% of image pixels to the valid class and the remaining pixels to the background class; (2) segment enhancement applying a morphological filter; (3) segment labeling using an 8-neighbor connected component criteria; (4) feature extraction for each segment (area, perimeter, compactness, elongation and invariant moments); and finally (5) classification using a linear decision function or a nearest-neighbor method. Working at night and acquiring images with mature fruits, the detection rates indicated that approximately 90% of visible fruit was detected successfully. In daylight, an 84% classification accuracy was declared and at least 20% of false detections. The vision system was designed to detect mature fruit but even under these circumstances there were certain problems when operating by day since the visible sky, the changing illumination conditions, and the sunlight intensity, which make a leaf appear brighter than a fruit if the leaf is illuminated directly by sun light, were highly perturbing factors. These were the main reasons to explain the false detections found, therefore the authors of this paper, in order to reduce false detections, recommended to use covers to eliminate background containing visible sky or the sun.

The European Eureka Project CITRUS ROBOT, involving both "Instituto Valenciano de

Investigaciones Agrarias"(Valencia, Spain) and CEMAGREF (Montpellier, France), investigated the robotic harvesting of oranges (Juste, 1991). This project considered not only the development of a vision system but other aspects such as an agronomy study including the definition of the working environment, fruit varieties and their manipulation, and also contemplated the development and control of a manipulator robot with its grasping device. For the vision system three methodologies based on color or spectral information were used. In the first approach, a B/W camera in conjunction with a red filter (630nm) and two synchronized stroboflashs were employed to obtain a uniformly illuminated scene to make the system as independent as possible from natural lighting conditions. Using a fast thresholding algorithm, 80% of the visible fruits were detected but a high number of false alarms were found. To improve the drawbacks of previous work a second approach was applied using an additional B/W camera. One camera used a red filter (630 nm) and the other a green filter (560 nm). Computing the ratio between the gray levels of both images, another image, which was independent of luminosity level, was obtained. Approximately 80% of the visible mature fruits were successfully detected and approximately 10% were false detections. Finally, in the third experiment, a color camera was used without artificial illumination. Each pixel, with its three RGB components, was considered as a pattern and was classified using a Bayesian method which was similar to the one presented by Slaughter and Harrell (1989). Success and false alarm rates of approximately 90% and 5%, respectively, were reported.

A vision system for harvesting melons was investigated under a collaborative study between Purdue University (USA) and Volcani Center (Israel). In the first attempt (Cardenas, 1991), a B/W camera was used to obtain intensity images of the melon crop. The computer algorithms to analyze these images could be divided in two main steps. In the first one, the purpose was to locate the melon on a horizontal plane, i.e to compute the X-Y coordinates corresponding to the center of the melon, and to estimate the melon size. This first stage performs an image enhancement, a segmentation by thresholding, a feature extraction and a hypothesis generation. Shape and texture parameters in the neighborhood of the hypothesized position were computed to obtain the final candidates. The second stage performed a knowledge-directed evaluation using rules which rejected noisy detections and multiple occurrences. If the second step was not employed, approximately 89% of existing melons were detected and high rates of false detections were found. However, using the knowledge-based rules, 84% and 10% rates were obtained, respectively. This means that the final evaluation eliminates most of the false detections.

The Hungarian Robotic Apple Harvester (AUFO robot), sponsored by the Hungarian-U.S.

Science and Technology Joint Fund, was developed from 1980 to 1989 and tested in 1990. This robot dealt with the harvesting of apples and performed automatic fruit detection using a stereo vision system which generated the 3D-dimensional position of each detected fruit. This vision system used two color cameras separated by a certain distance and oriented to converge over the same tree scene (Kassay, 1992). Once two images from the same scene but from different viewpoints were captured, a segmentation process was performed to obtain regions that should correspond to partial surfaces belonging to the apples. The regions were grouped and their geometric centers were obtained. For all possible pairs of segments that could be obtained by combining regions from both images, the three-dimensional position was computed. The technique used to compute this position was a simple triangulation algorithm divided into two steps. The first step gave the X-Y position using the projection on the X-Y horizontal plane that contained both cameras' optical axis, and the second step computed the heights or Z coordinates from each camera viewpoint. If the difference between these heights was less than 40 mm, then an object was considered to be present, otherwise the hypothesis was rejected. Only 41% of visible fruit was detected correctly and some false detections appeared. It was recognized that this approach had two main problems: (1) an exhaustive search was needed to check all the different possible segment pairs, which could also create some situations where some pairs were accepted as valid but in fact they were virtual fruit which actually did not exist; (2) since a tree has a natural configuration where occlusions are common, a triangulation system suffers the possibility of failing when trying to find two regions belonging to the same fruit captured from different viewpoints. This region correspondence problem, typical in stereo-vision systems, was magnified here, especially if the fruit was deep inside the tree, and seriously affected fruit detection ability.

A general vision system for the previously treated melon harvesting problem was presented by Dobrousin (1992). This vision system was divided into two subsystems: far-vision and near-vision modules. The far-vision subsystem used a B/W camera to locate the X-Y coordinates of the melon. The near-vision subsystem used a B/W camera and a laser light plane to extract the distance or Z coordinate, so that a picking arm could be guided towards the melon with precision. In this work, only the methodology used for the far-vision subsystem was shown. Several images were captured from the same scene, but the melons visibility changed because an air blowing system was used to avoid occlusion of the melons by leaves. These images were filtered, segmented by a histogram-based thresholding, cleaned by a morphological erosion operator and finally all the images were integrated by performing a logical OR operation. The resulting image was analyzed and some features (shape, area, size) were extracted from each segment. Finally, a rule-based classification was applied to obtain the valid fruits. Approximately 80% of the melons were detected and these gray level routines were integrated in a real-time pipeline system. The authors also proposed the use of infrared images to detect the differences of temperature that exist between leaves, soil and melons, which were specially significant in the afternoon when the melon temperature is lower than the temperature of the soil.

A description of the near-vision subsystem for the melon harvester robot was presented by Benady and Miles (1992). This system intended to compute the height Z of the melon center and refined the X,Y positions estimated by the far-vision subsystem. It used a plane of light to illuminate the scene and a converging B/W camera to register the intersection between the plane of light and the scene. This illumination system generated a straight luminous line when interacting with a flat surface, but when a curved surface was present, like a melon, a curved luminous line was recorded by the camera. The deformation of the curved line allowed the estimation of range information by a triangulation analysis. These deformed profiles were captured at regular spatial intervals while the system moved forward in the plantation. The profiles were analyzed using the Circular Hough Transform (CHT) to obtain an accumulator matrix indicating the most probable candidates for the center of the melon. To get the most probable candidates, the distribution around a pixel in the accumulation matrix was used instead of the absolute value in each cell of the matrix. For increasing the efficiency of the algorithm, some domain specific rules were used. These rules relied on the following parameters: the expected size, the shape, the relative distance to the soil, and the height value of the presumed fruit pixels that should belong either to melon patches or to leaves covering the fruit. All the fruits that were visually discernible were detected by the system, and no false detections occurred.

The vision system previously reported, related to the Spanish-French CITRUS ROBOT project, was unable to detect oranges during the initial stages of maturity, i.e. when the fruit was green. Plá (1993) proposed to solve this problem by using a different approach which relied on the detection of convex surfaces. Using flash-lamps and a B/W camera, an intensity image of the scene was obtained that should indicate convex surfaces or defined intensity gradients where a piece of fruit was present. This approach used the shape information and the intensity levels to detect spherical objects. The algorithm could be divided into two steps. The first stage computed the degree of convexity convoluting the Laplacian Of Gaussian function (LOG) with the intensity image. The resulting image was thresholded to consider only those pixels which had certain curvature and thereby reducing the computing time required for further steps. The goal of the second stage was to give a higher confidence level, based on shape information, about the existence of fruit. This stage consisted of fitting an ellipse to the initial image for all the points that passed through the convexity threshold. This fitting gave an error index indicating the accuracy of the fit in two directions, and finally this information was weighed and used in conjunction with the thresholded image to obtain the final segmented image. This system recognized oranges in the first stages of maturity with results of 75% and 8% of success and false detection rates, respectively. The false detections were mainly due to the presence of sky patches.

The Italian Agrobot robotic system for greenhouse operations like tomato harvesting (Buemi, 1995) was developed at CIRAA (Inter-University Center for Agricultural and Environmental Robots). The vision system used for this project was based on a color camera that supplied the HSI color components. Hue and Saturation histograms were employed to perform thresholding to segment the image. The 3-dimensional information was obtained by stereo-matching of two different images of the same scene. About 90% of ripe tomatoes were detected and the most frequent errors were due to occlusions.

There was an interesting research in the recognition of circular arcs which was applied to the detection of broken biscuits in sorting applications (Pla, 1996). In this work, the technique was also applied to the recognition of oranges in trees using a color camera. Since mature oranges were orange and leaves were green, the image had enough contrast so that an edge detection procedure could be applied and a clear contour image was obtained. The technique presented could be divided in two steps: an initial segmentation of contours generating groups of pixels arranged in constant curvature lines; and a second step where contour segments were grouped to obtain circle candidates and their parameters (radius, center and ratio of visible contours). The method worked very well when a good contour image was obtained, like in the biscuit application, but there were serious problems for the detection of fruit since the contour due to the occlusion of an orange by another orange or by a leaf generated false candidates.

Another vision system for the harvesting of oranges was investigated at the University College of London and A.I.D. center in Italy (Grasso and Recce, 1996). The robot they developed had two independent telescopic arms mounted on a common platform which was itself held by a large hydraulic arm. They used one small camera, fitted with a wide angle lens, inside each end-effector. Once the platform with the two small arms was positioned manually close to the canopy, the automatic harvesting system was invoked. Since two views from different angles were available, a stereo matching algorithm was used to obtain the position of each fruit. The RGB image data was thresholded to obtain a segmented binary image. After smoothing and edge detection, an adaptive algorithm was used to estimate the locations of centers by the Circular Hough Transform (CHT). During the process of approaching the fruit, the center of the orange was dynamically tracked, thus correcting the approaching course of the arm according to the movements of the fruit. The authors reported problems when oranges were partially occluded, when the segmentation algorithm found a leaf with sufficient orange color to be considered an orange, and other problems due to lighting conditions and wind speed. When the sky was very clear and the sun was low on the horizon, there was a high contrast in the image that exceeded the dynamic range of the CCD camera. In these cases, they reported that the segmentation technique could not work. Regarding the rate of correct and false detections, they reported 86% and 5% respectivelly, using a set of 50 images containing 673 oranges.

Recently, under the Spanish AGRIBOT project developed by the authors of this paper (Ceres et al., 1998), a different approach was presented to detect oranges, or any other fruit with spherical shape, regardless of fruit color or maturity stage (Jiménez et al., 1997, 1998 and 1999). They realized that most of the problems causing false detections, in systems using optical cameras, came from illumination variability and from some regions in the scene with similar segmentation results to those of real fruit. They proposed the use of a laser range-finder sensor able to obtain range images in a reliable way and based most of the recognition strategy on shape analysis. In fact, not only range information was captured but the intensity of the laser beam after reaching the range-finder sensor. This data was called reflectance and was used in conjunction with range data to analyze not only shape but optical object properties and object spatial distributions (Fig. 2). The recognition strategy focused on general spherical object recognition since oranges, apples, peaches or similar fruit can be modeled as spheres. The proposed approach was based on the generation of a set of primitives that were characteristic or very likely to belong to spherical objects. These primitives capture features lying on different areas of a sphere, therefore partial occlusions of the fruit could reduce the primitives captured but not the ability to detect the fruit. Four different primitives, called contour, crown, convex and reflectivity, were used to estimate the sphere parameters: 3D-position, radius and reflectivity. Tests were conducted using both artificial and natural tree configurations, reporting that approximately 80% of visible fruits were correctly detected and characterized. No false detections were found. The system was robust enough not to be influenced by fruit maturity stage, changes in illumination, shadows or the presence of background objects visible through the tree. The absence of false detections was a

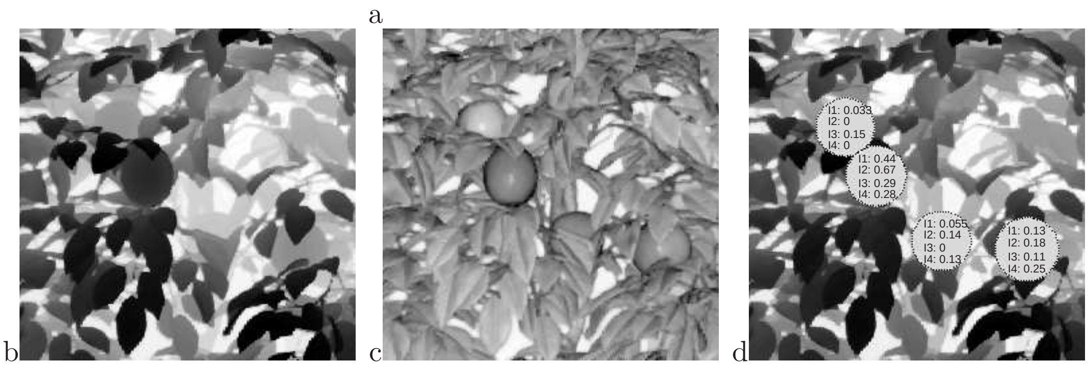

Figure 2: Captured images and detected fruit using the sensing system proposed within the Agribot project (Jimenez et al., 1998): (a) scene photo, (b) range image, (c) reflectance image and (d) detected fruit overlapped to range image.

proof of its robustness. The 3-D fruit location had a degree of precision around 10 mm, and the estimated radius and reflectivity could be used to perform a selective harvesting, i.e., harvesting only fruit of a certain size (using radius) or harvesting fruit at a certain maturity stage (using reflectivity). The current limitation of this system was the scanning speed and the processing time which required 20 and 60 seconds, respectively, for a 50×50 cm image running on a 150 MHz Pentium processor.

ANALYSIS AND DISCUSSION

The automatic detection systems previously described required three main steps: (1) image acquisition, (2) preprocessing for restoration or enhancement, and finally, (3) image analysis to detect the fruit. The studied solutions focused their attention mainly on the first and third stages, while the second received minor treatment. Table 1 presents a summary where the most important features of each approach are highlighted, including the research group, main references, fruit considered, sensors and accessories used, analysis method applied, ability to

Table 1: Summary of reported vision systems for detecting fruit on trees.

Research group & References	Fruit[1]	Sensors and accesories[2] (image type)	Analysis method[3] (algorithm details)	Detects green fruit	Correct-false detections[4]
U.Virginia (Parrish77)	Appl	B/W+F (Spectral)	Local (Thr+FExt+RCla)	No	N.R.
MAGALI (D'Esnon87)	Appl	Color (Spectral)	Local (Thr)	No	N.R.
(D'Esnon87, Rabatel88)	Appl	3 Color+ 3 F (Spectral)	Local (Ratio+Thr)	Yes	50%-high%
U.Florida and USDA (Slaughter87)	Oran	Color+ L (Spectral)	Local (Hue&Sat+LCla)	No	100%-N.R.
(Slaughter89, Harrell89)	Oran	Color (Spectral)	Local (RGB+ BCla)	No	100%-N.R.
U.Purdue (Whittaker87)	Toma	B/W (Intensity)	Shape (Contour+CHT)	Yes	68%-42%
A.I.D. (Levi88)	Oran	Color+ F+ L (Spectral)	Shape (Gradient+TMat.)	No	70%-N.R.
Sunkist and U.Calif. (Sites88)	Appl & Pech	B/W+ F+ L (Spectral)	Local (Thr+FExt+LCla)	No	84%-20%
AUFO (Kassay92)	Appl	2 Color (Spectral)	Local (Thr+stereo)	No	41%-N.R.
CITRUS (Juste91)	Oran	B/W+ F+ 2L (Spectral)	Local (Thr)	No	80%-high%
(Juste91)	Oran	2 B/W+ 2F+ 2L (Spectral)	Local (Ratio+Thr)	No	80%-10%
(Juste91)	Oran	Color (Spectral)	Local (RGB+ BCla)	No	90%-5%
(Pla93)	Oran	B/W+ L (Intensity)	Shape (Convx+ Thr&Fitting)	Yes	75%-8%
U.Purdue and Volcani (Cardenas91)	Meln	B/W (Intensity)	Local (Thr+CExt+RCla)	No	84%-10%
(Dobrousin92)	Meln	B/W+ Air (Intensity)	Local (Thr+CExt+RCla)	No	80%-N.R.
(Benady92)	Meln	Laser&B/W+ Air (Distance)	Shape (Profile+CHT+RCla)	Yes	100%-0%
CIRAA (Buemi95)	Toma	Color (Spectral)	Local (Hue&Sat+ Thr+ stereo)	No	90%-N.R
U.College-London (Grasso96)	Oran	2 Color (Spectral)	Local (Thr+ CHT+ stereo)	No	86%-5%
AGRIBOT (Jimenez97, 98, 99)	Oran & spheres	Laser Range finder (Distance & Spectral)	Shape & Local (4 primitives+ ParaEsti)	Yes	80%-0%

[1] Appl=Apples, Oran=Oranges, Toma=Tomatoes, Meln=Melons, Pech=Peaches.

[2] B/W= Black and White camera, Color= Color camera, F= Optic filter, L= Artificial light, Air=Air blower to move leaves.

[3] Thr=Thresholding segmentation, FExt=Feature extraction, TMat= Template Matching, LCla=Linear classifier, BCla= Bayessian classifier, RCla=Rule-based classifier, RGB= Red-Green-Blue feature space, Hue&Sat= Hue-Saturation feature space, CHT= Circular Hough Transform, Gradient=Local gradient image, Convx= Convexity image, Profile=Profile image, ParaEsti= Parameter Estimation by CHT and sphere fitting.

[4] N.R.=No Reported.

detect green fruit and performance in terms of correct and false detections. The following subsections give a detailed overview about the sensors and algorithms used emphasizing the limitations observed, the performance obtained and some recommendations or trends to apply in future research in this area.

Imaging sensors

The basic sensors employed in previous studies were B/W or color CCD cameras. Some of these used bandpass optical filters to increase contrast, as well as, artificial light to reduce shadows caused by sunlight. Others used laser technology to deduce range information from range-finders or by triangulation using a CCD camera. In the cases where the fruit had a fixed position, like in a melon crop, also an air blower was used to improve fruit visibility. In summary, it can be stated that studied digitalized images were basically from one of these three types:

1. *Intensity*
2. *Spectral*
3. *Range*

Intensity images were obtained using B/W cameras without the support of an optical filter. The illumination reflected at different wavelengths by the objects in the scene was captured by the camera CCD sensors as a brightness level independent of color. *Spectral* images were those which capture the reflected intensity at a specific wavelength. This could be achieved using a color camera which supplies the RGB or HSI components directly, or using a black and white camera with a colored filter to select a certain wavelength. *Range* images were those where each pixel corresponds to the the distance between the sensor and the sensed scene point. To obtain range images different techniques can be employed (Jarvis, 1983; Everett, 1995), but in this review only two techniques were reported: i) triangulation between a laser sheet and a optical camera, and ii) scanning a Local laser range-finder over the scene to get direct range and reflectance information without any processing.

Intensity and spectral images were appropriate when the brightness or color of the object to be recognized and the brightness or color of the background were different. In these cases segmentation was facilitated and the analysis algorithms could be very simple. Some problems affecting performance of vision systems using optical cameras to get intensity or spectral images were:

- *Shadows.* Due to sunlight illumination there were shadows and high contrast transitions that make applying direct recognition methods difficult. Fruit lighted by sun is ten times brighter than shadowed leaves and when the leaves were directly illuminated, they were four times brighter than shadowed fruit (Schertz, 1968). The use of artificial light could alleviate this problem but does not solve it completely (Sites, 1988). However, in some studies that used only color information without considering intensity values, most of the shadows and bright areas caused by direct sunlight did not create detection problems (Slaughter and Harrell, 1989).

- *No depth information.* Intensity or spectral images do not contain depth information. This means that localization was not complete and therefore it was required to move the arm along the visualization axis until the fruit was sensed. The use of stereoscopic techniques was an option to obtain the 3D-position but the correspondence problem in this environment was problematic since a tree scene consist of many similar objects and there was a high degree of occlusion which generated image segments without their corresponding pairs. Another alternative to obtain the depth information is the use of ultrasound or laser range sensors.

- *Confusing regions.* Some regions like patches of sky, sun or soil visible throughout the tree volume could generate areas in the intensity or spectral images with similar brightness or with an intensity distribution that generated convex segments that could be interpreted as valid fruit. In these cases, and to reduce false detections, authors recommended to work at night or to use protective coverings behind the tree to create a black background. In other cases, specifically when color images were used and a spectral analysis was performed, confusing regions were not prone to appear.

The use of range sensors is an alternative that does not suffer from the three above-mentioned problems. Shadows due to sunlight illumination do not influence the range data and images captured are independent of illumination levels. The patches of sky, sun or soil visible through leaves could be discarded from the valid range regions to be analyzed simply because the distance measured is outside the presumed working space or the strength of the laser signal is too weak. Furthermore, the 3D-position is directly obtained in the sensor spherical coordinate system which could be transformed to the robot Cartesian coordinate system to place a robotic gripper where the fruit was detected.

However, the total occlusion factor is probably the worst problem that a vision system could deal with. This problem affects not only systems based on optical cameras, but also range sensors. In some spanish orange crops it was estimated that 60% of the total existing fruit in a tree was totally hidden (Juste, 1994). Therefore, only 40% of partial occluded fruit is visible to a human operator. However, other authors in USA reported that the percentage of citrus visible outside the tree ranged from 70 to 100% (Schertz, 1968; Harrel, 1990). Therefore, these discrepancies in the visibility of citrus fruit indicate that fruiting conditions can vary significantly from region to region around the world, and depending on each case, the total occlusion problem could limit a harvester robot in a different way.

The total occlusion problem has been solved in the harvesting of melons using air blowers (Dobrousin, 1992; Benady, 1992), but this solution was only applicable to products like melons which were quite heavy, lying on the soil and with positions not shifted by air flow. This solution was not general, so other techniques should be used to reduce the occlusion. One already considered solution, for canopies with high occlusion, is the pruning of trees in conical shapes to increase the fructification over its periphery. It was estimated that using this strategy more than 75% of total fruit would be visible (Juste, 1994). The harvesting of fruit following a descending order instead of an ascending operation results in an additional increase of visibility of 7-8%. This fact is easily explained because the removal of fruit from upper portion of the trees allows branches to move upward, clearing leaves away from the non-harvested lower portion of the trees.

Image Analysis Methods

Regarding the recognition processes, the revised techniques can be categorized in two main groups according to the features they use to perform the fruit detection:

1. *Local-based*
2. *Shape-based*

Techniques based on *local* features, use the value or values associated with each image pixel to decide whether a pixel belongs to a fruit or to the background. These discriminant values are the pixel intensity and color components in spectral images (RGB, HSI or specific isolated wavelengths captured using an optical filter). Recognition methods based on local data had simple and fast algorithms and therefore were highly interesting when real-time applications were considered. However, in general, these methods were sensitive to varying conditions, as a result, the algorithms need frequent re-training or re-tuning for some thresholds to continue performing well.

Techniques based on *shape* information analyze region convexities or edge distributions forming circular arcs generated when profiles or contours from spherical objects were extracted. These methods were more general than local methods since the former ones could be applied to detect fruit independent of its color. However, applying shape-based methods when intensity or spectral images were sensed is not a recommended approach since illumination transitions, shadows and patches of sky or sun visible through the tree might produce regions with convex distribution or circular contours that would be misinterpreted. Research that look for certain shape configurations using intensity or spectral images (Whittaker, 1987; Levi, 1988; Pla, 1993), were all characterized by significant false detection rates. Nevertheless, shape methods operating over range images (Benady, 1992; Jiménez, 1997-98-99), provided robust solutions. One disadvantage of shape-based methods is that they require special hardware to work in real-time.

Performance

In this paper, three main sensing methods have been reviewed: intensity, spectral, and range. Two main analysis strategies were applied: local-based and shape-based. This review presents several fruit detection systems which derive from the different approaches that could be obtained combining the possible types of sensing and analysis stages:

- *Intensity/local* (Cardenas, 1991; Dobrousin, 1992).
- *Intensity/shape* (Whittaker, 1987; Plá, 1993).
- *Spectral/local* (Parrish, 1977; D'Esnon, 1987; Slaughter, 1987-89; Harrell, 1989-90; Sites, 1988; Kassay, 1992; Juste, 1991; Buemi, 1995; Grasso, 1996).
- *Spectral/shape* (Levi, 1988).
- *Distance/shape* (Benady, 1992; Jiménez, 1997-98-99).

The majority of research used local analysis based on spectral images. The best results using this approach indicate correct detection rates of almost 100%. Nevertheless this strategy was restricted to the detection of fruit with a different color to that of the background and could present false detections if this condition is not satisfied. Shape-based methods applied to intensity or spectral images had the advantage of detecting fruit regardless of its color, but false detections were quite frequent. Methods using range images and shape-based strategies obtained good results since the systems were capable of dealing with different-colored fruit, did not generate false alarms and the percentage of correct detections were above 80%.

CONCLUSIONS

This paper has reviewed different computer vision approaches for the detection of fruit on a tree. A description of each reported approach in this field was presented, emphasizing the kind of sensor used, the analysis strategy applied to detect the fruit, and the performance in terms of correct and false detections as well as its sensibility to changes in color.

Three different types of images were used in the reviewed works: intensity, spectral or range images. They were obtained by means of CCD color or black and white cameras, artificial illumination and optical filters, in most of the cases, or using laser range sensors that supplied the distance information. In terms of the detection algorithms employed, all approaches used either local or shape-based analysis. The performance reported depends on the global strategy chosen and the particular features of each considered application.

The type of image captured was a highly influential factor that constrained the system performance since further processing steps were going to be based on the information obtained from that image. Therefore, sensor selection is a crucial task that should be accomplished to appropriately allow the capture of images containing stable and discriminant information that facilitate software analysis as much as possible.

Total occlusion was a problem that seriously affected the practical application of the systems reviewed, regardless of the image and analysis types used. Therefore, alternatives to reduce total occlusion must be investigated, including more pruning studies for external fructification, or the use of non-traditional sensors capturing absorption properties that differentiate fruits from leaves. This approach should generate images where leaves were not visible and where only fruit were displayed, allowing for an easy detection process.

Even in the case of an ideal tree scene registration, the rocking motion or oscillation of fruits in a tree due to wind streams, which is normal when operating in external environments, creates a time-variant fruit position error that troubles the harvesting process. Since fruit detection and grasping/detaching processes are both required to implement a harvesting robot, if the fruit detection process only gives the 3-D fruit position, which could be only valid at a certain time, the subsequent detaching process would be incapable of finding the fruit at the referenced position. Accordingly, a special gripper and end-effector control should be used to deal with this position variability using passive mechanical designs and active sensing strategies to track, in a close-loop manner, the local fruit displacements as the robot arm extends.

References

Benady M. and Miles G.E. 1992. Locating melons for robotic harvesting using structured light. *Paper ASAE No.:92-7021.*

Buemi F., Massa M., and Sandini G. 1995. Agrobot: a robotic system for greenhouse operations. *4th Workshop on robotics in Agriculture, IARP, Tolouse*, pages 172–184.

Cardenas M., Hetzroni A., and Miles G.E. 1991. Machine vision to locate melons and guide robotic harvesting. *Paper ASAE No. 91-7006.*

Ceres R., Pons J.L., Jiménez A.R., Martín J.M. and Calderón L. 1998. Agribot: A robot for aided fruit harvesting. *Industrial Robot*, 25 (5).

Dobrousin Y. and Y. Edan. 1992. Real-time image processing for robotic melon harvesting. *Paper ASAE No.:92-3515.*

Duda R.O. and P.E. Hart. 1972. Use of the hough transformation to detect lines and curves in pictures. *Comunications of the ACM*, 15(1):11–15.

D'Esnon A.G., G. Rabatel, and R. Pellenc. 1987. Magali: A self-propeled robot to pick apples. *ASAE paper 87-1037, ASAE, St Joseph, MI 49085-9659.*

Everett H.R. 1995. *Sensors for mobile robots. Theory and applications.* A.K. Peters, Ltd. Wellesley, Massachusetts.

Gaffney J.J. 1969. Reflectance properties of citrus fruit. *Transactions of the ASAE*, 16(2):310–314.

Grasso G. and M. Recce. 1996. Scene analysis for an orange picking robot. *Proc. Intern. Congress for Computer Technology in Agriculture (ICCTA'96).*

Harrell R.C. 1987. Economic analysis of robotic citrus harvesting in florida. *Transactions of the ASAE*, pages 298–304.

Harrell R.C., D.C. Slaughter, and P.D. Adsit. 1989. A fruit-tracking system for robotic harvesting. *Machine Vision and Aplications*, 2:69–80.

Harrell R.C., P.D. Adsit, R.D. Mumlla, and D.C. Slaughter. 1990. Robotic picking of citrus. *Robotica*, 8:269–278.

Illingworth J. and J.Kittler. 1988. A survey of the hough transform. *Computer Vision, Graphics and Image Processing*, 44:87–116.

Jarvis R.A. 1983. A laser time-of-flight range scanner for robotic vision. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 5(5):505–512.

Jiménez A.R. , R. Ceres, and J.L. Pons. 1997. Shape-based methods for fruit recognition and localization using a laser range-finder. *Bio-Robotics-97. International Workshop on robotics and automated machinery for bio-productions*, Gandía, pages 155–160.

Jiménez A.R. 1998. *Sistema de reconocimiento y localización de objetos cuasi-esféricos por telemetría láser. Aplicación a la detección automática de frutos para el robot Agribot.* Tesis Doctoral. Universidad Complutense de Madrid.

Jiménez A.R., R. Ceres, and J.L. Pons. 1999. Automatic fruit recognition: A survey and new results using range-attenuation images. *Pattern Recognition*, 32:1719–1736.

Juste F., I. Fornes, F. Plá, E. Molto, and F. Blay. 1991. Primeros resultados en campo de un prototipo de brazo robotizado para la recolección de cítricos. *CIMA91, 23 Conf. Int. Maquinaria Agrícola, Zaragoza*, pages 433–440.

Juste F. 1994. *Memoria de actividades 1993/1994.* Instituto Valenciano Investigaciones Agrarias. IVIA.

Kassay L. 1992. Hungarian robotic apple harvester. *Paper ASAE No. 92-7042 St. Joseph, MI 49085*, pages 1–14.

Levi P., R. Falla, and R. Pappalardo. 1988. Image controled robotics applied to citrus fruit harvesting. *Procedures ROVISEC-VII, Zurich.*

Parrish E. and A.K. Goksel. 1977. Pictorial pattern recognition applied to fruit harvesting. *Transactions of the ASAE*, 20:822–827.

Plá F., F. Juste, and F. Ferri. 1993. Feature extraction of spherical objects in image analisys: an application to robotic citrus harvesting. *Computers and Electronics in Agriculture*, 8:57–72.

Plá F. 1996. Recognition of partial circular shapes from segmented contours. *CVGIP: Computer Vision and Image Understanding*, 63(2):334–343.

Rabatel G. 1988. A vision sistem for magali, the fruit picking robot. *Paper 88293, AGENG88, Int. Conf. Agricultural Engineering, Paris.*

Schertz C.E. and G.K. Brown. 1968. Basic considerations in mechanizing citrus harvest. *Transactions of the ASAE*, pages 343–346.

Sites A. and M.J. Delwiche. 1988. Computer vision to locate fruit on a tree. *ASAE paper*

85-3039, ASAE, St Joseph, MI 49085.

Slaughter D. and R.C. Harrel. 1987. Color vision in robotic fruit harvesting. *Transactions of the ASAE*, 30(4):1144–1148.

Slaughter D. and R.C. Harrel. 1989. Discriminating fruit for robotic harvest using color in natural outdoor scenes. *Transactions of the ASAE*, 32(2):757–763.

Whittaker, Miles, Mitchell, and Gaultney. 1987. Fruit location in a partially ocluded image. *Transactions of the ASAE, 30*, pages 591–597.