

A vision system based on a laser range-finder applied to robotic fruit harvesting

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>

1 Abstract

This paper describes a laser-based computer vision system used for automatic fruit recognition. It is based on an infrared laser range-finder sensor that provides range and reflectance images and is designed to detect spherical objects in non-structured environments. Image analysis algorithms integrate both range and reflectance information to generate four characteristic primitives which give evidence of the existence of spherical objects. The output of this vision system includes the 3-D position, radius and surface reflectivity of each spherical object. It has been applied to the AGRIBOT orange harvesting robot, obtaining good fruit detection rates and unlikely false detections.

2 keywords

laser-scanner, harvesting, reflectance, range

3 Introduction

Today, any production system intending to be competitive by means of reducing production costs and improving product quality, has to consider the use of automatic techniques. Computer vision techniques are being used to analyze images of agricultural products for grading purposes under controlled illumination with successful results, automating real-time sorting tasks previously done by human operators. However, the automatic recognition of fruit for agricultural harvesting is a problem not satisfactorily resolved yet, due to the complexity of this environment. Therefore, the harvesting of delicate fruit such as oranges, peaches or apples is now being performed by hand. Some of the most representative attempts to automate the detection and location of fruit come from the research developed for these robots: MAGALI [19](Rabatel 1988), AUFO [13](Kassay 1992) and CITRUS [17](Plá et al. 1993). These papers, as well as others related to computer vision methods for locating fruit on trees were surveyed by Jiménez [10](Jiménez et al. 2000b).

The recognition of apples was considered in the MAGALI and AUFO projects, as well as, in other research works done at the Universities of Virginia [16](Parrish and Goksel 1977) and California [21](Sites and Delwiche 1988). The Univ. of Florida collaborating with the USDA center [22](Slaughter and Harrel 1989), the Italian AID Institute [14](Levi et al. 1988) and the Spanish/French CITRUS project [17](Plá et al. 1993), developed methods for the detection of oranges. Tomatoes were the target at the University of Purdue [23](Whittaker et al. 1987) and in the Italian Agrobot project [2](Buemi et al. 1995). On the other hand, the harvesting of melons was studied at the University of Purdue and the Israeli Volcani center [1](Benady and Miles 1992).

The above mentioned solutions, used color or black & white CCD cameras as the main sensing devices, in some cases using artificial lighting sources and optical filters to reduce the natural shadows and improve the contrast between fruits and background. The only case where a different sensor was used was for the recognition of melons [1](Benady and Miles 1992), where they used a structured light triangulation method to obtain range profiles.

Regarding the analysis method, most of them were based on local properties of each pixel in the image [19][13][16][21] [22][2](Rabatel 1988, Kassay 1992, Parrish and Goksel 1977, Sites and Delwiche 1988, Slaughter and Harrel 1989, Buemi et al. 1995), performing for each pixel a classical feature classification based on a feature space built up with RGB/HSI or intensity components. This approach was suitable when fruits and leaves had different spectral properties, usually being a straightforward and reliable technique for recognizing mature fruit since leaves are green-colored objects while mature fruit has that color component attenuated. However, these color-based techniques were not capable of coping with non-mature fruit which is an important problem given that some fruit varieties, and under some conditions, must be harvested when the peel is still green. Additionally, color-based approaches typically produced a significant number of false detections caused basically by non-leaf background objects such as patches of sun, soil or sky.

Other approaches used shape information for analysis [17][14][23][1](Plá et al. 1993, Levi et al. 1988, Whittaker et al. 1987, Benady and Miles 1992). They perceived the global relationships among pixels in the image, which could be arranged forming convex surfaces due to the gradual decrease in intensity, when going from the center of each fruit towards its periphery. The circular arrangement of contour edge pixels or profiles were also detected using circular Hough transform strategies, but in some cases false detections were reported due to the circular contours that edges of leaves generate when a high-pass filter was applied [23](Whittaker et al. 1987).

Both intensity/color pixel-based and shape-based analysis methods were appropriate strategies for the recognition of fruits, but some problems arose from the variability of the sensed image itself when using CCD cameras, which are very sensitive to changes in sunlight intensity as well as shadows produced by the leaves. Since leaves lighted by the sun can be brighter than fruits in the shadow, the intensity images obtained are likely to generate difficulties for the analysis algorithms. Using color images the features are more stable, but there are also problems with background objects with a color similar to the fruits. Additional to these lighting problems, partial and total occlusion are, without any doubt, the worst problems because the visibility area can be reduced drastically affecting significantly the recognition performance.

It is known that the percentage of partially visible fruit compared to the total existing fruit in a tree goes from 40% [12](Juste and Sevilla 1991b) to 70% [20](Schertz and Brown 1968) depending mainly on fruiting and viewing conditions. This fact obviously limits the practical application of these systems unless specially pruned tree configurations were obtained to increase the fruit visibility. If percentages are related to the total number of visible fruit, one of the best reported detection rates comes from the CITRUS project [11][17](Juste et al. 1991a, Plá et al. 1993), which shows correct detections rates above 90% and 5% for false detections. These results are valid for situations where the fruit to be detected has a different color than the background, however when the vision system uses BW cameras trying to detect both mature and unripe fruit, the best results obtained are 75% and 8% for correct and false detection.

In a different direction, the Spanish AGRIBOT project where the authors are involved [3][4](Ceres et al. 1996, Ceres et al. 1998), tried to solve the recognition of oranges following two approaches: automatic and manual. The automatic mode uses the vision system we present in this paper to detect most of the visible fruit without false detections. After that, the manual mode is activated to locate the remaining fruits. In this mode, the operator visually detects fruits, handles a joystick to point to each fruit using a laser range-finder, and finally, their three-dimensional locations are computed.

As some authors have pointed out, shape-based methods using a reliable range sensor could provide a good strategy for object recognition [6](Jain and Flynn 1993). Following this strategy, to implement the above described automatic detection system for the AGRIBOT robot, we propose to base the analysis process on shape, reflectivity and fruit spatial distribution, and use as a basic sensor an infrared laser range-finder providing range and reflectance data. The following sections

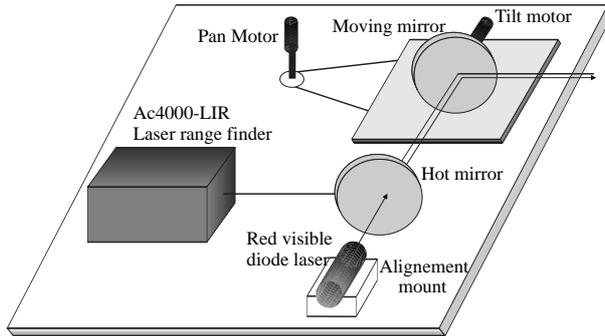


Figure 1: Scanning system for the laser range-finder sensor.

describe the different topics tackled to implement this laser-based machine vision system.

4 Laser range-finder sensor

4.1 Sensor selection and configuration

Although there are several techniques to obtain depth information like stereo vision, dynamic vision, focus, texture, shape from shading, structured lighting, ultrasound or Moiré fringes; in our case, considering some desirable technical specifications for the sensed image, the best choice is the use of a telemetric system. The specifications that our sensor must cope with are: generation of absolute range and reflectance data, range up to 4 meters, 1 mm standard deviation consistency, 10 mm range accuracy, high measurement rate ($>10\text{kHz}$) and coaxial emission/reception path to avoid missing data in the images.

However the search for a commercially available sensor satisfying those specifications is difficult, especially when the system must be also eye-safe, i.e., the power of the laser beam must be limited. It has been demonstrated that there is a trade-off between the different parameters that characterize a laser range-finder [7](Jiménez 1998):

$$M = \frac{D}{\sigma_D \sqrt{T \cdot F_i}} \quad (1)$$

where D is the maximum range the system can measure, σ_D is the noise or range measurement consistency, T is the measurement period, F_i is the laser output power and M is an efficiency factor value which is higher when the telemeter global specifications are better. This relationship (Eq. 1) means that using a certain technology M to design a laser range-finder, we can sacrifice some of the parameters (for example F_i) to obtain better performance in the remaining (D , σ_D and T). This also means that is not possible to improve all the technical specifications as a whole, unless a higher technology were used.

The technical specifications we are looking for were analysed within a large range of commercially available laser sensors [7](Jiménez 1998), detecting a technological limit ($M = 500$) that none of the studied sensors were able to cross. After an extensive search, only one sensor coped with the required specifications. This is the commercially available laser range-finder AccuRange-4000-LIR, which operates with a phase-shift measuring technique ($D = 15$ m; $\sigma_D = 2.5$ mm; $T = 2 \cdot 10^{-5}$ s; $F_i = 8$ mW; $M = 474$). This sensor was configured to allow for the scanning of a scene by means of a moving mirror driven by two DC motors in pan and tilt directions (see Fig. 1). Since the laser wavelength is 780 nm (infrared) this beam is invisible, therefore to permit a human operator to indicate the area of interest to be scanned, a visible diode laser, with a cross generator, was aligned parallel to the invisible beam using a hot mirror.

The AccuRange-4000-LIR sensor provides range data and a signal that is proportional to the strength of the optical signal received (reflectance). Therefore, for each scanned scene two images are obtained: range and reflectance (see Fig. 2). Range images acquired with this sensor are not

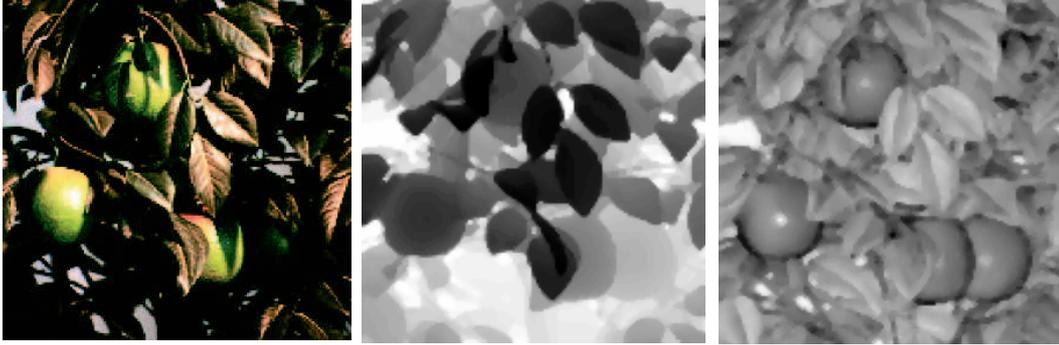


Figure 2: Left to right: Scene picture, range image and reflectance image.

affected by sun illumination changes and shadows, because the laser itself is the lighting source used to illuminate the scene. Additionally, if non-specular surfaces are present, the range image is quite reliable, which allows an analysis based on the topographical shape of the range map. On the other hand, the reflectance data does not have a clear meaning and that is why we have a special interest to model this data and study the internal and external parameters affecting the registered reflectance value.

4.2 Reflectance model

A mathematical model which connects reflectance with other factors that affect its value, can be used to interpret much better reflectance images. One of these factors is the distance, D . It means that reflectance and distance are related by an equation, so it allow us, in different ways, to integrate both sources of information to facilitate scene analysis. For instance, we could use the reflectance map to make a shape analysis compatible to range data, or we could derive optical properties from the surface we are measuring using reflectance and range data in conjunction.

Basically, the reflectance value registered by the laser range-finder is closely related to the beam intensity received throughout the sensor's reception window after suffering absorption, reflection and dispersion phenomenon when hitting the target surface, plus some final transformations inside the sensor such as focusing, filtering, amplification, digitalization and averaging. Most of these phenomenon are well studied, so we incorporated them in our model, including some other factors specific to this laser range-finder. Finally, we will validate the model comparing it to experimental data. These are the key properties included in the reflectance model:

- *Second Lambert law.* The radiant intensity $I[w/sr]$ detected at a certain observation angle, θ , over a lambertian surface, is modulated by a cosine, i.e.:

$$I = I_{\theta=0} \cos \theta \quad (2)$$

where θ is the angle between the surface normal and the observation axis.

- *Signal attenuation by square of distance:* D^2 .
- *Target surface reflectivity:* ρ . Reflectivity measures the ratio between the incident light and the total reflected light. It is a parameter that depends on surface properties and on the wavelength, λ , of the illuminating source. In our case, ρ is constant because the laser has a fixed wavelength (780 nm).
- *Specific sensor features.* Such as the optic arrangement (β ; optical attenuation), area of reception window ($A = 2827 \text{ mm}^2$), laser output power ($F_i = 8 \text{ mW}$) and logarithmic amplification of received electric signal.

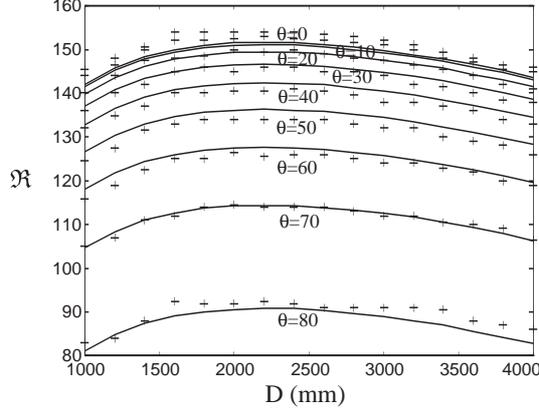


Figure 3: Least square fitting between reflectance model and experimental data.

Using the second lambert law (Eq. 2), and expressing $I_{\theta=0}$ in terms of the radiant flux $F_i[w]$ and the diffuse reflectivity coefficient ρ , we obtain

$$I = \frac{F_i}{\pi} \rho \cos \theta \quad (3)$$

and considering the solid angle, Ω , covered from the sensor's reception window to the measured point at a distance D , we can derive the following expression (Eq. 4) for the radiant flux, F_c , captured by the photoelectric detectors inside the sensor.

$$F_c = \beta I \Omega = \left(\frac{\beta A F_i}{\pi} \right) \frac{\rho \cos \theta}{D^2} \quad (4)$$

Since the electromagnetic energy is transformed into an electrical signal and then this signal is amplified logarithmically, we can state an initial reflectance model for the AccuRange-4000-LIR laser range-finder:

$$\begin{aligned} \mathfrak{R} &= cte_1 \log(cte \cdot F_c) = \\ &cte_1 \log\left(cte_2 \frac{A F_i \rho \cos \theta}{\pi D^2}\right) \end{aligned} \quad (5)$$

which has been validated using empirical data (Fig. 3), finding that good least square fits were obtained when $cte_1 = 80.9$ and cte_2 had a linear dependency with D due to defocussing effects. The slight discrepancies between the experimental data and the model, specially when θ is close to 0, appear because our model only considers lambertian surfaces. However, this approximation is valid for most of the natural surfaces.

If $\alpha(D)$ includes the variable optic attenuation, the photodiode sensibility and the signal preamplification, we can obtain this reflectance model for lambertian surfaces:

$$\mathfrak{R} = 80.9 \log \left(\alpha(D) \frac{A F_i \rho \cos \theta}{\pi D^2} \right) \quad (6)$$

where \mathfrak{R} and D are the reflectance and distance generated by the sensor, respectively, θ the angle between the normal to the object surface and the laser beam axis, ρ is the diffuse reflectance of the object surface for a 780 nm wavelength laser source and A , F_i are the area of the reception optics and the incident radiant flux, respectively. This model shows the relationship between \mathfrak{R} , D , θ and ρ , and therefore, we will use it to integrate both range and reflectance images at a low-level stage, and to obtain surface reflectivity information that can be used to facilitate the segmentation process as will be shown in section 4.

5 Adaptive image smoothing

Range information supplied by the sensor is mainly degraded with gaussian noise. The amplitude of this degradation has been modeled (Eq. 7) showing there is a dependency with the reflectance value, \mathfrak{R} , the sampling rate, T , and the maximum range the sensor is configured to measure, D_{max} . The derivation of this consistency model is presented in detail in appendix A.

$$\sigma_D = \sqrt{\frac{8.37 \cdot 10^{-3}}{T \cdot 10^{\mathfrak{R}/80.9}} + \left(\frac{D_{max}}{9 \cdot 10^7 T}\right)^2} + 1 \quad (7)$$

In order to detect scene features or primitives that require the computation of local first or second order derivatives in a robust way, is necessary to restore the range map. The ideal result of the restoration is a range image with no fluctuations or rugosity over the continuous surfaces. Also, it should have minimum bias error or oversmoothing at the discontinuities, compared to the noise-free image. Therefore, the restoration method must supply smoothness and fidelity. We performed a restoration evaluation between different filters (DW-MTM, MAS, MMSE, Median, Wavelet, Sigma, Gaussian and others), and did not find outstanding quality results in the restored range images when considering both smoothness and fidelity [9](Jiménez et al. 2000). So, we proposed the design of a new adaptive filter coping with both factors.

The proposed filter, named 3σ -MPF, is based on the 3σ -fidelity test, which is defined as follows

$$3\sigma - test = \begin{cases} 1 & \text{If } \forall(x, y) \in v_{ij} \\ & |g(x, y) - (ax + by + c)| \leq 3\sigma \\ 0 & \text{Otherwise} \end{cases} \quad (8)$$

where v_{ij} is the neighborhood of pixel (i, j) , the x, y pairs are the neighboring pixels, g is the map to be restored and (a, b, c) are the parameter defining a plane fitted to the v_{ij} region.

This criteria is used to determine two crucial aspects when designing adaptive filters: (1) the determination of the kernel size that it is moved across the image, and (2) the detection of a discontinuity. To determine the kernel size a multi-resolution plane fitting approach starting from the highest window and dropping to smaller when the 3σ -fidelity test fails (residuals are high) is applied. If a plane is fitted and it fulfills the 3σ -fidelity test, then it is assumed to be on a continuous region and the kernel mean is considered as the restored range estimation. When a minimum kernel size is reached without finding a successful plane fit, then it is assumed to be on a discontinuity and a MTM (Modified Tream Mean) algorithm is applied.

Fig. 4 presents the pseudocode of the algorithm for the 3σ -MPF filter. For computational reasons kernel sizes were restricted to windows of 7×7 , 5×5 and 3×3 pixels. The standard deviation values, needed to compute the residuals, were estimated from the reflectance image using equation 7. This filter reduces most of the noise, smoothing surfaces while preserving the discontinuities.

A sample of the results obtained with the 3σ -MPF filter is shown in Fig. 5. Profiles in the filtered range image are quite smooth and free of noise allowing stable derivative computations. Additionally, discontinuities are preserved which is a very important fact to maintain range fidelity, limit position errors and maintain significant shapes.

6 Recognition strategy

Fruits such as oranges, apples or peaches can be approximately modeled as spheres, so the fruit recognition problem can be stated as a more general spherical object recognition problem. The simplicity of a sphere should allow us to perform its recognition by straightforward template matching techniques, at least in structured environments, but the presence of leaves that partially occlude fruits and have circular contour and profiles, makes a direct template matching which is not an appropriate technique for this recognition problem. The reason is that correlation peaks that belong to spheres or background are not clearly separated.

```

FOR each image pixel g(i,j)
   $\sigma$ =noise_estimation(g,(i,j));
  kernel_size=7x7;
  REPEAT
    a,b,c=least_square_plane_fitting(g,(i,j),kernel_size);
    3 $\sigma$ _fidelity_test=3 $\sigma$ _fidelity(g,(i,j),kernel_size,(a,b,c), $\sigma$ );
    IF 3 $\sigma$ _fidelity_test==FALSE
      kernel_size=kernel_size-2x2;
    UNTIL (3 $\sigma$ _fidelity_test==TRUE OR kernel_size<3x3);
    IF 3 $\sigma$ _fidelity_test==TRUE // continuous region found
      f(i,j)=c;
    ELSE // discontinuity or impulse detected
      ref=Median(g,(i,j),3x3,4_connectivity);
      f(i,j)=3 $\sigma$ _range_mean(g,(i,j),3x3,8_connectivity,|g(x,y)-ref|<3 $\sigma$ )
    END
  END

```

Figure 4: 3σ -MPF adaptive filter pseudocode.

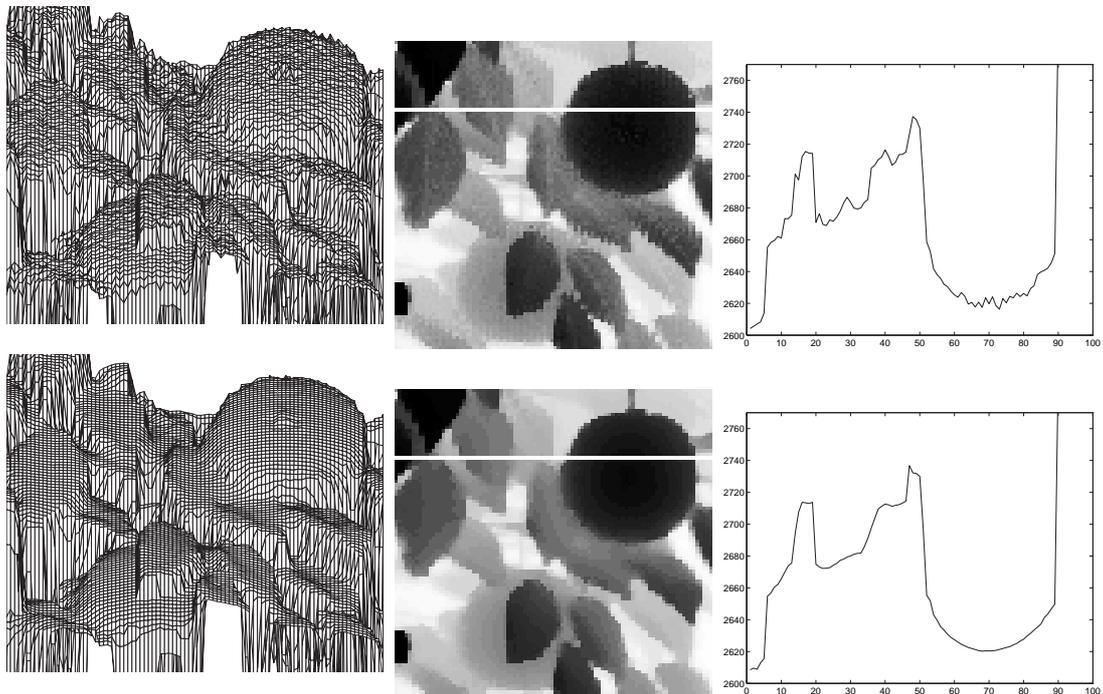


Figure 5: 1st row: unfiltered range image from an orange tree scene. 2nd row: restored case using the adaptive 3σ -MPF filter. Each row, from left to right: range map, intensity coded range map and range profile coming from the horizontal white line.

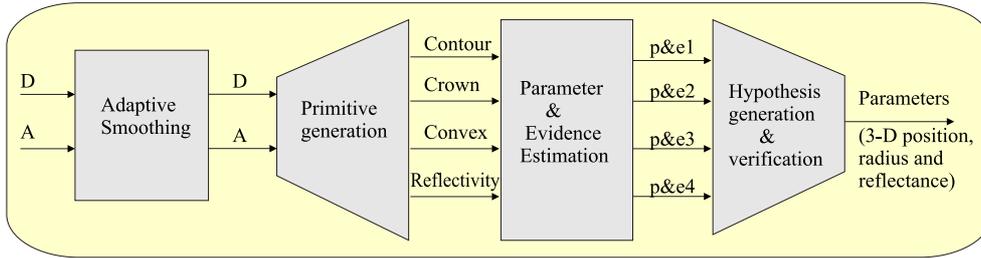


Figure 6: Recognition strategy block diagram.

The approach we propose is based on the generation of a set of primitives that are characteristic or very likely to belong to spherical objects. We extract them and accumulate evidence about the existence of a sphere using a hypothesis generation approach. Some of these primitives are point-primitives obtained analyzing a small neighborhood surrounding each pixel, and also there are region-primitives extracted from large smooth segments. From each region-primitive, we could deduce the sphere parameters with a degree of confidence, on the other hand, point-primitives do not generate direct evidence about the presence of a sphere, but its global arrangement forming circular arcs does. The local definition of these point-primitives allow them to cope with most of the occlusion problems. Although occlusion will reduce the evidence, the detection capacity remains unaltered while there are sufficient primitives to generate consistent hypotheses.

A block diagram with the main stages of our recognition strategy are depicted in Fig. 6. The stages for primitive extraction, parameter estimation and hypothesis generation/verification are briefly explained in the following subsections.

6.1 Primitive generation

The primitives we considered most useful to recognize spherical objects are four: contour pixels, crown pixels, convex regions and reflectivity-based regions. These primitives were selected because its ability to cover the whole surface of a sphere with little overlapping, and therefore, since they detect different areas of the sphere, these primitives are appropriate in cases with high surface occlusion. A primitive extraction example for the scene of Fig.5 is displayed in Fig.7.

Contour primitives are obtained after a pixel classification process where range and reflectance images are used to extract four features which measure the gradients and the local variation of gradients in both range and amplitude images. These four features allow us to classify pixels in five classes: sphere edges, crease edges, other edges, intensity step edges and smooth pixels. Sphere edge is a step discontinuity at a sphere boundary, and they are a subset of the classical step edges obtained with a Sobel or Canny edge extractor. Sphere edges will be our contour primitives and are ideal for our objectives because they include sphere boundaries but avoid circular arcs generated by leave boundaries.

The second primitive was called a *crown* point. We observed that there is an optimum slope value in range images that maximizes the likelihood of a pixel to belong to the surface of a sphere and at the same time minimizing the probability of getting pixels from plane-like objects. If a sphere is present, and a thresholding is performed with this slope value, a circle of crown points is obtained over the periphery of the sphere. Since flat surfaces with an orientation equal to the optimum slope will generate many crown pixels, the curvature of the surface is also used by accepting only pixels with correct slope values in both range and reflectance images, and at the same time, their gradient directions are checked to see if they are compatible.

The *convex* and *reflectivity* regions, are obtained using the smooth pixels found in the pixel classification step described above, after applying a connected component classical algorithm. Only large regions with mean curvature and standard deviations close to those we are looking for, are finally considered as convex regional primitives. To decide whether a region should be considered a reflectivity primitive or not, a thresholding operation is used. The discriminant reflectivity threshold is obtained by an off-line learning process. If fruit and the background have separated

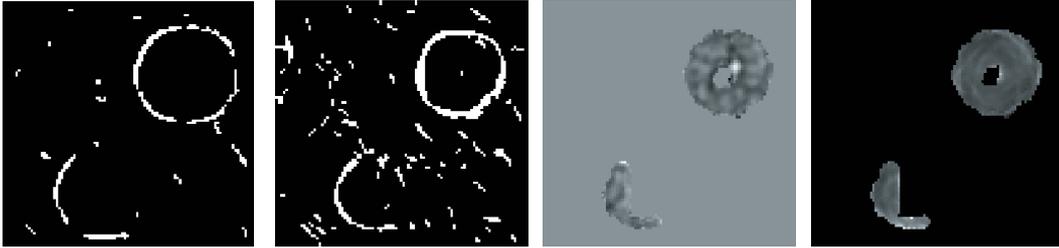


Figure 7: Left to right: Contour, crown, convex and reflectance primitives.

modes in a reflectivity histogram, the system uses the optimum discriminant threshold value.

6.2 Parameter and evidence estimation

The purpose of this stage is to estimate the sphere parameters (3-D position, radius and reflectivity) and the degree of confidence we have over that estimation. Each primitive is used together with the D and \mathcal{R} images to obtain the parameters. Contour and crown primitive images consist of certain fragmented circular arcs surrounded by noisy pixels, so a robust estimation technique is necessary and therefore the circular Hough transform was employed [24](Zhang 1997). To improve the performance, the dimensionality of the parameter space is reduced using information about the expected radius of the fruit, the number of votes per pixel is reduced using edge direction information, and peak detection in the parameter space improves using the back-transformation method [5](Illingworth and Kittler 1988). The evidence for these estimated parameters depends on the ratio between the number of votes and the maximum possible number of votes. Convex and reflectance primitives are clean-labeled segments and don't need any special robust method. They are used to initialize a minimum squared sphere fitting process over the range image to obtain the parameters and a confidence measure.

6.3 Hypothesis generation & Verification

In this step the four partial hypothesis are integrated to generate the final hypothesis. Compatible parameters are clustered and the evidences are added so weak candidates based on one kind of primitive are reinforced by other compatible weak detections to generate a more confident hypothesis. After the integration, those hypotheses not having sufficient evidence value are rejected (10% of maximum evidence). Finally, to eliminate possible false detections, a rule-based rejection step is applied over the accepted candidates to take out incoherent detections. The rules include the testing of valid 3-D positions, radius and the existence of holes in the fruit or volumetric intersections between pairs of fruits.

7 Results and conclusion

Two evaluation strategies were applied using simulated and empirical data. In the first case we analyzed the sensibility of the evidence value, that measures the confidence of the hypothesis, adding some perturbances to an ideal range/reflectance image containing a sphere. These disturbances were different quantities of gaussian noise (σ : 0-6 mm), impulsive noise (0-5%), occlusion (0-80%), eccentricity of spheres (1.0-1.2) and spatial resolution (1-6 mm/pixel). Over these ranges of degradation, the system performance was affected by reducing the evidence value of the spherical hypothesis, but the spherical detection capability remained. Therefore, the software algorithms demonstrated robustness against degradation.

Empirical tests were performed using the laser scanner and an artificial orange tree with green and orange fruits placed with different degrees of visibility (see Fig. 8). Approximately 80-90% of visible fruits were correctly detected, although diminishing when the spatial resolution is low



Figure 8: Empirical test setup.

or the evidence threshold increases. No false detections were found if the evidence threshold was above 0.15, but this is not a constraint because the maximum evidence is 4.0 (1.0 for each of the partial hypothesis). See Fig. 9 to see the change of correct and false detections when the evidence threshold varies. The dashed lines indicate the performance obtained if only one primitive at a time were used. From this we can see that using the four methods the different primitives complement each other to obtain better overall results.

Precision in the 3-D localization of the fruit is approximately 10 mm, and the estimation mean error of the fruit radius is below 5 mm, although these values are highly dependent on the scanning spatial resolution. Fig. 10 shows an example of the recognition and parameter estimation process. The final hypotheses are indicated over the range image overlapping a circular gray disk at a position and radius that correspond to the estimation. We can see how the four oranges on the scene are correctly detected.

Probably, the main limitation of the current system is the acquisition speed (1 kHz), which means that almost 25 seconds is needed to get a 160x160 image. The speed could be increased but the accuracy would be decreased as equation 17 shows, which would affect the reliability of spherical fruit registration. Parallelization of the scanning and the analysis process could reduce the initial acquisition time and should allow for picking cycles around 6 seconds, which is not too far off from the estimated picking cycle of 2 seconds for our two manipulator AGRIBOT robot system [4](Ceres et al. 1998).

Despite the cited technical limitations, this approach is robust enough not to be influenced by maturity stages of the fruit, changes in illumination, sunlight shadows or visible background objects behind the tree. The absence of false detections and the good detection rate is the best proof of its robustness. Reliable 3-D positions and radius values are obtained as well as a reflectivity value which could be used to perform a selective harvesting, picking only the fruits with the desired maturity stage.

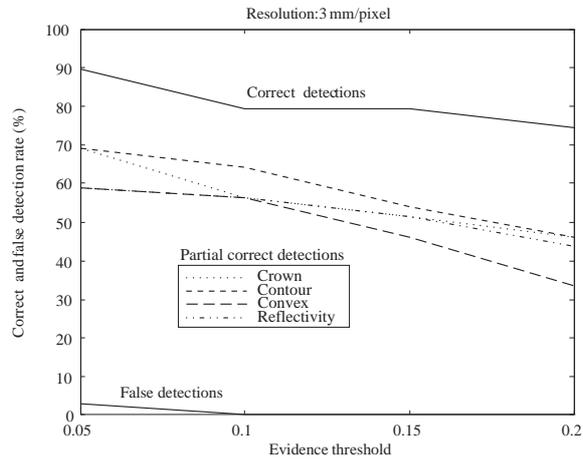


Figure 9: Correct and false detection rates (res.: 3 mm/pixel).



Figure 10: Fruit recognition example. From left to right: scene photo, range, reflectance images and detections found.

8 Appendix: Range consistency model

This appendix presents a range consistency model which shows the standard deviation of range readings, σ_D , as a function of some factors. Depending on the measuring conditions, a typical consistency value above 1 mm and below 8 mm is typical for the AR-4000-LIR sensor.

There are several noise sources for an optical measuring system like our telemeter. This noise comes from the photodetector (photonic noise), the quantification stages, the emitting laser diode, and thermal and amplification processes. In our case, the dominant noise sources are the photonic and the quantification noise, $\sigma_{D(ph)}$ and $\sigma_{D(q)}$ respectively. Other minor noise sources are included under the symbol $\sigma_{D(other)}$. Considering that these sources are not correlated, we assume the total noise that affects range data, σ_D , can be modeled by the equation

$$\sigma_D = \sqrt{\sigma_{D(ph)}^2 + \sigma_{D(q)}^2 + \sigma_{D(other)}^2} \quad (9)$$

Considering the photonic noise, if we represent the photon energy by E , it can be represented using the following equation

$$E = \frac{hc}{\lambda} \quad (10)$$

where λ is the laser wavelength, h is the Planck constant ($6,62 \times 10^{-34} \text{ J} \cdot \text{s}$) and c is the speed of light.

The number of photons, n , received in the laser range-finder photodiode during the interval of time T , is

$$n = \frac{F_c T}{E} = \frac{\lambda T}{hc} F_c \quad (11)$$

where F_c is the radiant flux captured by the laser telemeter (see Eq. 4).

The number of photoelectrons, n_{phe} , generated at a photodiode with a quantum efficiency η in the interval T , is $n_{phe} = \eta n$. Assuming a photoemission process following a Poisson distribution [18](Pratt 1991), the photoemission standard deviation, σ_{phe} , equals the square root of the number of photoelectrons emitted, n_{phe} . Estimating the photodetector signal to noise ratio, $SNR_{(ph)}$, as the ratio between the electrons n_{phe} emitted in the interval T , and the photoemission noise σ_{phe} , the following expression is derived:

$$SNR_{(ph)} = \frac{n_{phe}}{\sigma_{phe}} = \frac{n_{phe}}{\sqrt{n_{phe}}} = \sqrt{n_{phe}} = \sqrt{\eta \frac{\lambda T}{hc} F_c} \quad (12)$$

Substituting F_c in previous equation gives

$$SNR_{(ph)} = \sqrt{\frac{\lambda T}{hc} \eta \frac{\beta A F_i \rho \cos \theta}{\pi D^2}} \quad (13)$$

Identifying terms with the reflectance model (Eq. 6), we have

$$SNR_{(ph)} = \sqrt{\frac{\lambda}{hc} T 10^{\mathfrak{R}/80.9}} = 20 \times 10^8 \sqrt{T 10^{\mathfrak{R}/80.9}} \quad (14)$$

Range standard deviation, σ_D , for a telemetric system based on phase shift, is inversely proportional to $SNR_{(ph)}$ [15](Nitzan et al. 1977). Then, standard deviations due to the photonic noise can be expressed as:

$$\sigma_{D(ph)} \propto \frac{1}{\sqrt{T 10^{A/80.9}}} \quad (15)$$

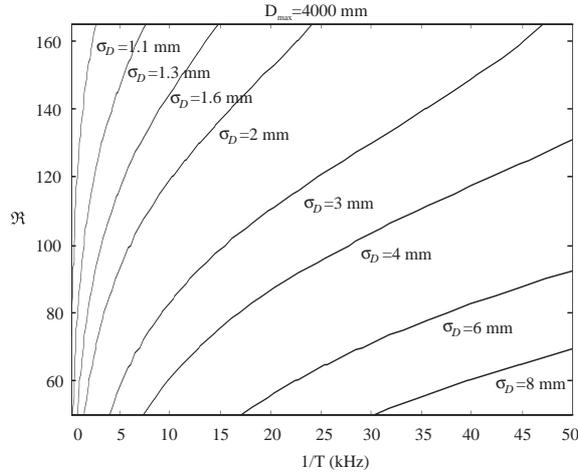


Figure 11: Range noise estimation by the Consistency Model.

Considering the second source of noise, the quantification noise $\sigma_{D(q)}$, the laser range-finder designers assure that its sensor follows this expression:

$$\sigma_{D(q)} = \frac{D_{max}}{9 \times 10^7 T} \quad (16)$$

where D_{max} is a programmable factor for setting the maximum distance the sensor can measure (expressed in mm), and T is the sampling interval in seconds.

Finally, grouping other noise sources under the term $\sigma_{D(other)}$, whose magnitude has been obtained experimentally having a constant value of 1 mm, we can present the final consistency model for the AC-4000-LIR sensor, which shows there are three factors influencing the range measurement fidelity: \mathfrak{R} , T y D_{max} :

$$\sigma_D = \sqrt{\frac{8.37 \cdot 10^{-3}}{T \cdot 10^{\mathfrak{R}/80.9}} + \left(\frac{D_{max}}{9 \cdot 10^7 T}\right)^2} + 1 \quad (17)$$

Fig. 11 gives a quantitative idea of the noise level that the range measurements suffer under different operating conditions (changes in reflectance \mathfrak{R} and sample interval T). It can be deduced that even in the best case, a 1 mm additive noise is expected, with more than 8 mm at short sampling rates and reflectances around 50 units.

References

- [1] Benady M and Miles G (1992) Locating melons for robotic harvesting using structured light. Paper ASAE No.:92-7021
- [2] Buemi F, Massa M, and Sandini G (1995) Agrobot: a robotic system for greenhouse operations. 4th Workshop on robotics in Agriculture, IARP, Toulouse, pages 172–184
- [3] Ceres R, Pons J, Jiménez A, Martín J, and Calderón L (1996) Agribot: A robot for aided fruit harvesting. AgENG96-Paper 96A-107
- [4] Ceres R, Pons J, Jiménez A, Martín J, and Calderón L (1998) Design and implementation of a fruit harvesting robot (Agribot). *Industrial Robot*, 25 (5)
- [5] Illingworth J and Kittler J (1988) A survey of the Hough transform. *Computer Vision, Graphics and Image Processing*, 44:87–116
- [6] Jain A and Flynn P (1993) *Three-dimensional Object Recognition Systems*. Elsevier

- [7] Jiménez A (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 (Ph.D.). Universidad Complutense de Madrid
- [8] Jiménez A, Ceres R, and Pons J (1999) Automatic fruit recognition: A survey and new results using range-attenuation images. *Pattern Recognition*, 32: 1719-1736
- [9] Jiménez A, Ceres R, and Pons J (2000) A new adaptive filter and a quality evaluation index for image restoration. To appear in *Pattern Recognition*
- [10] Jiménez A, Ceres R, and Pons J (2000) A survey of computer vision methods for locating fruit on trees. Submitted to *Transactions of the ASAE*
- [11] Juste F, Fornes F, Plá F, Molto E, and Blay F (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
- [12] Juste F and Sevilla F (1991) Citrus: A european project to study the robotic harvesting of oranges. *Proceedings of the 3rd International Symposium on Fruit, Nut and Vegetable Harvesting Mechanization. Denmark-Sweden-Norway*
- [13] Kassay L (1992) Hungarian robotic apple harvester. Paper ASAE No. 92-7042 St. Joseph, MI 49085, pages 1-14
- [14] Levi P, Falla R, and Pappalardo R (1988) Image controlled robotics applied to citrus fruit harvesting. *Procedures ROVISEC-VII, Zurich*
- [15] Nitzan D, Brain A, and Duda R (1997) The measurement and use of registered reflectance and range data in scene analysis. *Proc. IEEE*, 65:206-220
- [16] Parrish E and Goksel A (1997) Pictorial pattern recognition applied to fruit harvesting. *Transactions of the ASAE*, 20:822-827
- [17] Plá F, Juste F, and Ferri F (1993) Feature extraction of spherical objects in image analysis: an application to robotic citrus harvesting. *Computers and Electronics in Agriculture*, 8:57-72
- [18] Pratt W (1991) *Digital Image processing*. Wiley
- [19] Rabatel G (1988) A vision system for magali, the fruit picking robot. Paper 88293, *AGENG88, Int. Conf. Agricultural Engineering, Paris*
- [20] Schertz C.E. and Brown G.K. (1968) Basic considerations in mechanizing citrus harvest. *Transactions of the ASAE*, pages 343-346
- [21] Sites A and Delwiche M (1988) Computer vision to locate fruit on a tree. ASAE paper 85-3039, ASAE, St Joseph, MI 49085
- [22] Slaughter D and Harrel R (1989) Discriminating fruit for robotic harvest using color in natural outdoor scenes. *Transactions of the ASAE*, 32(2):757-763
- [23] Whittaker, Miles, Mitchell, and Gaultney (1987) Fruit location in a partially occluded image. *Transactions of the ASAE*, 30, pages 591-597
- [24] Zhang Z (1997) Parameter estimation techniques: a tutorial with application to conic fitting. *Image and Vision Computing*, 15:59-76

9 Biographies

Antonio R. Jiménez graduated in Physics, Computer Science branch (Universidad Complutense of Madrid, June 1991). He received the Ph. D. degree in physics sciences from the Complutense University of Madrid, October 1998. From 1991 to 1993, he worked in industrial laser applications at CETEMA (Technological Center of Madrid), Spain. From 1994, he is working as research assistant at the Instituto de Automática Industrial, CSIC, Spain. His current research interests include computer vision applications, pattern recognition, range images, shape-based image analysis and automatic harvesting.

Ramón Ceres graduated in Physics (electronic) from Universidad Complutense of Madrid in 1971 and received the doctoral degree in 1978. He works at the Instituto de Automática Industrial (IAI-CSIC). As a researcher, Dr Ceres has worked on sensor systems for different fields such as continuous process control, machine tool, agriculture, robotics and disable people. On these topics has more than seventy papers and congress presentations, and has several patents in industrial use. At present Dr Ceres is the Spanish delegate for the IMT (Brite-Euram) Committee and Deputy Scientific Director of the IAI.

José L. Pons graduated as Mechanical Engineering (Universidad de Navarra, April 1992). He received a M. Sc. degree in Information Technologies for Production (Universidad Politécnica de Madrid, January 1995). He received the Ph. D. degree in physics sciences from the Complutense University of Madrid, December 1996. Dr. Pons is currently at the Instituto de Automática Industrial, CSIC, where has been working from 1993. His current research interests include non-traditional sensor-actuation technologies, development of new technologies and miniature applications.