Available online at www.sciencedirect.com
ScienceDirect
Journal of Food Engineering 87 (2008) 371–379
JOURNAL OF FOOD ENGINEERING
www.elsevier.com/locate/jfoodeng

Olive classification according to external damage using image analysis

M.T. Riquelme *, P. Barreiro, M. Ruiz-Altisent, C. Valero

Physical Properties Laboratory and Advanced Technologies in Agrofood (LPF-TAG), Rural Engineering Department, ETSI Agrónomos, Polytechnic University of Madrid, Av. Complutense s/n 28040 Madrid, Spain

Received 7 September 2007; received in revised form 20 December 2007; accepted 20 December 2007
Available online 1 January 2008

Abstract

The external appearance of an olive's skin is the most decisive factor in determining its quality as a fruit. This work tries to establish a hierarchical model based on the features extracted from images of olives reflecting their external defects. Seven commercial categories of olives, established by product experts, were used: undamaged olives, mussel-scale or 'serpeta', hail-damaged or 'granizo', mill or 'rehús', wrinkled olive or 'agostado', purple olive and undefined-damage or 'molestado'. The original images were processed using segmentation, colour parameters and morphological features of the defects and the whole fruits. The application of three consecutive discriminant analyses resulted in the correct classification of 97% and 75% of olives during calibration and validation, respectively. However the correct classification percentages vary greatly depending on the categories, ranging 80–100% during calibration and 38–100% during validation.

Keywords: Table olives; External damages; Artificial vision; Sorting fruit; Image processing

1. Introduction

The presence of skin damage in olives is the most decisive factor in determining their external quality as fruit. Traditionally, olives have been sorted manually. Only recently have some research institutes and manufactures been looking for techniques allowing for automatic classification. Such research has worked towards the development of prototype devices for fruit classification based either on artificial vision (Diaz et al., 2000, 2004; Mateos et al., 2005) or on sorting techniques like fruit rebound (Barreiro et al., 2003).

Recently, Brosnan and Sun (2004) presented an extensive review of the different image processing techniques for food products, which are increasingly used technologies. Many applications have been developed using artificial vision as a technique for fruit classification: peaches (Cordero et al., 2006), citrus (Blasco et al., 2007; Kondo et al., 2000), cherries (Rosenberger et al., 2004; Uthaisombut, 1996) and especially apples (Cheng et al., 2003; Kavdir and Guyer, 2004; Mehl et al., 2004). However, there is a lack of studies focused on discriminating a large number of external defects in olives based on the commercial categories established by product experts, as proposed in this paper. In all refereed works the following steps are used for image processing: acquisition, pre-processing, segmentation, measurement (extraction features) and interpretation. The segmentation process is an essential step in image analysis. Flawed segmentation can cause interpretation errors. According to Du and Sun (2004), there are several segmentation techniques available for food quality evaluation, which may be grouped into four different categories: histogram, region, edge and classification-based. The selection of one method or the other depends on the case. This work examines whether the combination of colour parameters and histogram features extracted from an image might enhance classification.

Leemans and Destain (2004) used a hierarchical grading method and *k*-means clustering for a real-time grading system, achieving correct apple classification with a success rate of 73%. Also Kleynen et al. (2005) introduced a defect

* Corresponding author.
E-mail address: mt.riquelme@upm.es (M.T. Riquelme).

0260-8774/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jfoodeng.2007.12.018

Nomenclature

A maximum (r,g,b)
r,g,b normalized RGB coordinates
B minimum (r,g,b)
Bi selected B of gray level "i" for the blue component
Centroid position (x,y) of the center of mass of the region
ConvexArea number of pixels in the polygon which circumscribe the region, consisting of the tangents of its border.
Eccentricity ratio of the distance between the foci of the ellipse and its major axis length
EquivDiameter diameter of a circle with the same area as the region ($4 \cdot \text{area}/\pi$).
FilledArea actual number of pixels in the region with all holes filled in
Gi selected G of gray level "i" for the green component
H hue, colour type
Hi selected H of gray level "i" for the hue component
MajorAxisLength length in pixels of the major axis of the ellipse
MinorAxisLength length in pixels of the minor axis of the ellipse
Number of objects in our case, number of objects in the image (the labeling only identifies one element, although there may be various regions of interest
Orientation angle (degrees) between the x-axis and the major axis of the ellipse
Perimeter number of pixels in the length of the outline of the defect's region
Solidity proportion of the pixels in the smallest polygon which contain the regions, which are also in the region (area/convexarea)

Ratio Perimeter/Area
Ratio 'Area of defect'/'Area of olive'.
Ri selected R of gray level "i" for the red component
Roundness perimeter2/$4 \cdot \pi \cdot$ area
S saturation, colour intensity
Si selected S of gray level "i" for the saturation component

Subtraction 'Area of olive' from 'Area of defect'.
T threshold
V value, colour brightness
Vi selected V of gray level "i" for the value component

segmentation procedure consisting in a pixel classification algorithm based on the Bayes' theorem and non-parametric models of the sound and defective tissues, where segmentation of russet defects and colour transition areas of skin were problematic. Recently, Blasco et al. (2007) proposed region-based segmentation based on unsupervised techniques for detecting the most common external citrus defects; in this segmentation algorithm, the contrast between different areas in the image becomes more important than the individual pixel colour. Some researchers combine several segmentation techniques for defect detection, like for example Bennedsen and Peterson (2005): three threshold segmentation routines and another based on artificial neural networks and principal components.

For histogram-based methods, thresholding can be applied locally to calculate different thresholds for each pixel within a neighborhood (Niblack, Sauvola, etc.) or to globally calculate one single threshold for the whole image (isodata, entropy, etc.). However, histogram-based thresholding is still the most referenced among segmentation methods (Barreiro et al., 2008; Unay and Gosselin, 2006) based on the method described in Otsu (1979). Researchers continue studying and comparing the different techniques for the detection of external defects on fruits; Unay and Gosselin (2006) obtained better performance by applying thresholding for the segmentation of apples when compared to local segmentation.

Usually colour is the main focus in artificial vision systems, as it is an important quality parameter in fruits and vegetables. Different colour coordinate spaces are used for the description of colour. Frequently CIE $L^*a^*b^*$ is used, as it shows the objective colour with fidelity. However, in the artificial vision field, colour is defined by RGB coordinates or HSI coordinates, the latter being equivalent to cylindrical coordinates (CSI) of $L^*a^*b^*$. The HSV model is very similar to the HSI colour model. The main difference between the two is the calculation used to produce brightness values. In the HSI model, a pixel's brightness (I) is derived from the mean of its three (R, G and B) colour values. In the HSV model, a pixel's brightness (V) is determined by the mean of the minimum and maximum of its three colour values. However, colour is one of the most uneven aspects when dealing with olives.

There are commercial systems that allow the packer to sort olives automatically according to their external appearance, based on optical properties and image analysis. These systems use different colour spaces, although the expectations of the producer, the packer and the manufacturer of the classification device do not match exactly the results obtained in terms of the percentage of correct classification. Standards are needed as well as reference methods to verify the performance of such systems.

The present work tries to obtain classification functions to sort olives into different quality classes, according to the

external flaws on olives of the Spanish "Manzanilla sevillana" variety. The final methodology could be useful to evaluate the performance of commercial systems, but it is not aimed at replacing them.

2. Materials and methods

2.1. Fruit

A representative sample of the 260 olives of the "Manzanilla sevillana" variety was selected to define the mathematical model (Table 1). Olives were previously classified by a set of product experts (Fig. 1) at several processing factories using common terms. The Alvarado et al. (2001), FAO (1987), Trapero and Blanco (2001) and USDA (1967), provide definitions for the commercial categories available for olives:

- *Undamaged olives:* Fruits harvested during the ripeness cycle, before seasonal colour changes, and with their varietal characteristics. Free of defects.
- '*Serpeta*' *or mussel-scale* (*Lepidosaphes ulmi* L.): This kind of cochineal is very widespread throughout the world and causes damage to pome fruits and olives. The damaged olives show dry and elongated scars on the epidermis.
- *'Granizo' or hail-damaged:* This meteorological phenomenon can cause serious damage to the olive's epidermis due to strong impacts (round marks).
- *'Rehús' or mill:* specific term for rejectable olives.
- *Wrinkled olive or 'agostado':* There are several external factors (temperature, humidity, etc.) which can influence the dehydration of green olives; the fruits have shrunk and present a dry appearance.
- *Purple olive:* Colour is an indicator of fruit ripeness. In the case of purple olives is the colour for full-ripeness, which is present in the skin as well as in the pulp. This commercial group includes olives whose colour diverges greatly from the commercial characteristics required for the fruit.

Table 1
Damage types identified in olives by product experts at processing companies

Damage types (commercial categories)	Calibration	Validation
Undamaged olives	16	14
'Serpeta' or mussel-scale	5	6
'Granizo' or hail-damaged	7	5
'Rehús' or mill	5	18
Wrinkled olive or 'agostado'	16	3
Purple olive	27	41
'Molestado1' or undefined-damage	20	29
'Molestado2' or undefined-damage	27	21
Total	123	137

The set was divided for the calibration and validation test.

Fig. 1. Damage types identified in olives by expert technicians at processing companies: (a) undamaged olives, (b) 'serpeta', (c) 'granizo', (d) 'rehús', (e) 'molestado1', (f) wrinkled, (g) purple olive and (h) 'molestado2'.

- *'Molestado' or undefined-damage:* All external defects that do not affect the olive pulp are included in this commercial-group: direct hit, bruises, scratches, superficial marks of different shapes and sizes, etc. These damages are typical of poor handling during harvest. This category has been divided in two groups to facilitate classification: 'molestado1' (round shape bruises) and 'molestado2' (elongated shape bruises). Mateos et al. (2005) in a study with olive images, also identify two types of 'molestado'.

The calibration set was built with those samples which only had one category of defect in the skin ($N = 123$ for calibration), while the rest of olives with a mixture of defects were used for the validation set ($N = 137$). This work aims to study a wide population containing all of the defects defined above. However, some categories are represented by a small population ('serpeta' and 'granizo'), variable, depending on harvest and environmental conditions. As a result, it is difficult to obtain a larger sample in some cases.

2.2. Reference data

Reference tests include destructive and non-destructive experiments: weight, fruit size (height and equatorial diameter per fruit), fruit colour as determined by a Minolta Spectrophotometer (CIE $L^*a^*b^*$), and by using an RGB image camera. RGB values were transformed into HSV using an algorithm of Matlab (version 7.0; Math Works, Inc., USA), as the online sorting device works in this colour space. The different damages were characterized for each fruit following a previous study of olives (Barreiro et al., 2003).

2.3. Imaging system

Olive images in RGB were acquired under static conditions with a single-sensor (one CCD: charge-coupled device) colour camera (model JVC-TK 1270E; JVC, Victor Company of Japan). The camera spatial resolution is 0.1 mm^2/pixel.

The colour camera was installed on a mobile column. The sample was placed inside an hemispheric integrating sphere of white expanded polystyrene foam, in order to generate an homogeneous diffuse lighting to get rid of shadows during image acquisition. Samples were lit with fluorescent lights with high colour rendering index (CRI) (model TL/95; Phillips, Royal Philips Electronics of the Netherlands), that is, providing colour reproduction equivalent to 95% of that afforded by sunlight).

2.4. Image processing

To facilitate the acquisition process, all original RGB images contained two olives. It was necessary to perform three consecutive segmentations over the original images: the first segmentation to leave out the background, then to identify one olive per photo, and finally a segmentation of external defects inside each fruit area.

Some researchers, like Bennedsen and Peterson (2005) agree to use a combination of different threshold segmentation routines for better extraction of fruit and defect features.

The algorithms used were developed with the image processing toolbox, version 6, of Matlab (Mathworks, 2005).

2.4.1. Image pre-processing

First of all, the fruit area was separated from the background with the algorithm of Otsu (1979), which is a nonparametric and unsupervised method of automatic threshold selection for picture segmentation, based on statistical and space information of histograms (Gaussians distributions). Using the Matlab function 'graythresh', a discriminant criterion makes it possible to obtain the optimal threshold of an image that maximizes the group variable (σ^2).

The procedure is very simple: when the threshold (T) is applied, the image is converted automatically into binary values (black and white), where '1' means background and '0' foreground (object):

$$g(x,y) = \begin{Bmatrix} 1 \leftrightarrow f(x,y) < T \\ 0 \leftrightarrow f(x,y) > T \end{Bmatrix} \quad (1)$$

Thus, the thresholding level obtained after applying Otsu's algorithm is a normalized intensity value that lies in the range $[0,1]$. The thresholding was performed on the three colour components together, as the Matlab routines analyzed them as one single matrix. The images were converted into binary using automatic thresholding.

It was found that a simple thresholding was not enough, as the binary images showed two big elements (olive areas), and in some occasions, several little elements (due to shadows under poor light) in the background part of the images. Thus, a filter was applied to remove small objects from the image. A morphological opening operator was used with a disk-shaped structuring element considering a six-pixel radius (0.6 mm in the image). Thus the objects with a radius lesser than the selected value were eliminated. Afterwards another morphological operator (object filling) was used to fill in the possible holes presented in the segmented binary images.

2.4.2. Defect extraction

In order to segment the defects from the sound olive area, images were readjusted for each RGB channel so as to fulfil the intensity values 0–255; this increases the contrast of the output image. However the Matlab function 'imadjust' works with intensity values between 0 and 1 (double), therefore the images to adjust were previously transformed to this scale. In order to select the upper and lower limits for this process all the olive images were studied; in our case, the specific ranges set – so as to include 99% of pixel values – resulted in the lower (Red 0.6, Green 0.2, Blue 0) and upper (Red 0.7, Green 0.9, Blue 1) intensity levels. Values below low-in and above high-in are clipped.

The rest of the algorithms applied for defect extraction were similar to those used for whole fruits: thresholding, eliminating small elements, filling in holes (disk-shape with a radius of four-pixels (0.4 mm in the image), and labelling objects. The labeling operation was similar to that described for isolating olives in the image.

2.4.3. Characterization of fruit images

After segmentation, the olives were characterized by different features taken from each fruit from all filtered images (both whole-fruits and the defect extracted).

- *Parameters extracted referring to the olives:* The colour coordinates analyzed were RGB and HSV. The RGB image is transformed into HSV with the 'rgb2hsv' routine of the image processing toolbox, version 6, of Matlab (Mathworks, 2005), which uses the following equations:

$$\begin{array}{lll} \text{If} & A = r; & H = \frac{\left(\frac{g-b}{A-B}\right)}{6} \\ \text{If} & A = g; & H - \frac{\left(2+\frac{b-r}{A-B}\right)}{6} \\ \text{If} & A = g; & H = \frac{\left(4+\frac{r-g}{A-B}\right)}{6} \end{array} \left| \begin{array}{l} S = A - B \\ V = A \\ A = \text{maximum} \\ B = \text{minimum} \end{array} \right. \quad (2)$$

  Each colour parameter was separated: Red, Green, Blue, Hue, Saturation and Value. When the corresponding histograms are studied, the intensity level distribution contributes to the general description of the images. Histograms were filtered by a moving average in order to decrease noise: a window with size of 15 data were used.

The following descriptive statistics were used to characterize each histogram: Mean, Median, Maximum, Mode, Kurtosis, Skewness, Quartile, and Quintile.

- Parameters extracted referring to defects:
  Several geometrical parameters were defined and quantified, defined in the nomenclature table.

2.5. Statistical analysis

A covariance matrix for each of the histogram matrices (123 olives × 256 gray levels) corresponding with the histogram of each colour parameter (R, G, B, H, S and V) was studied. Note that the diagonal of the matrix corresponds to the variance of the corresponding gray intensity levels. Fig. 2 shows an example of the visualization of the covariance matrix.

Significant differences between categories of olives were determined using variance analysis (one-way ANOVA) was performed by Statistica (version 6.1; StatSoft, Tulsa, OK, USA). Next, a Fisher's least significant difference (LSD) test was used to determine the significant differences between group means in an analysis of variance ($p \leqslant 0.05$). This test was applied to all parameters in order to reduce the huge amount of parameters analyzed.

The significant variables were introduced into a forward stepwise discriminant analysis (DA) to obtain the classification functions in successive steps. The *a priori* probability to belong to each group was set as equal in all cases. The canonical function obtained is a linear combination of discriminating attributes, being the sum of raw canonical coefficients multiplied by the coefficients for each function (StatSoft, 2007).

Three different discriminant analyses (DA) were performed in this work to segregate among defects ($N = 123$ calibration set). The validation of the model was performed with a sample of $N = 137$. Thus, the first DA classified the olives into two categories: global defects and local defects. For each group obtained, two successive DA were required in order to segregate all defects.

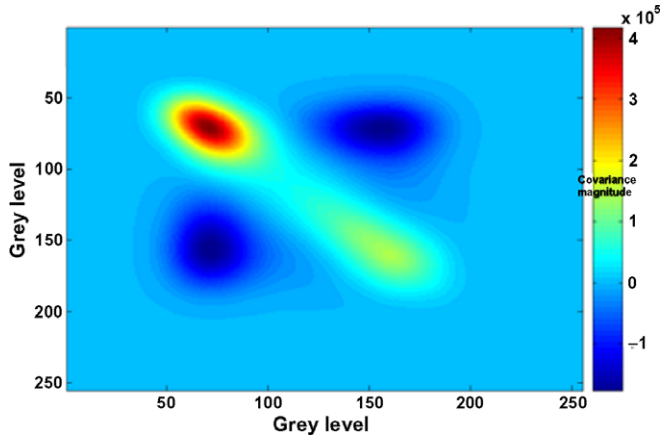

Fig. 2. Covariance matrix cov (x_i) for the red channel. Areas where covariance is higher (red) were used to extract two gray levels (x,y) in each graph as discriminatory variables. (For interpretation of the references to colour the reader is referred to the web version of this paper.)

3. Results and discussion

3.1. Characterization of fruits

Table 2 shows an ANOVA for the reference values. Letters correspond to the LSD test. Data was arranged according to fruit weight. Three groups of olives were found according to the weights: purple olives (5.0 ± 1.0 g/fruit), wrinkled olives (1.9 ± 0.6 g/fruit), and the rest (3.5–4.4 g/fruit). Also, the wrinkled olives have a smaller caliber, as expected in keeping with their weight. With the characterization of the olives, reference tables were created for directed users of the image analysis technology in order to clarify commercial agreements. As a confidential result of this study, such tables with the image, colour, size and firmness of each defect type are now in the hands of manufactures and producers (Riquelme et al., 2004).

Regarding colour change with ripeness it was verified that: darker, more red and less yellow olives presented lower L^*, higher a^* and higher b^* values, respectively. Thus, the colour parameters correctly separate the global defects with a purple colour of skin ('rehús', wrinkled, and purple) from the rest.

Although some catalogued defects are well segregated by the characterization parameters (Table 2), this is not sufficient to separate defects in the different commercially established categories. We thus used the characterization parameters extracted from the images and their histograms to go on with the classification process by image analysis.

3.2. Image analysis: calibration and validation

As a first approach, the shape and distribution of the histograms were enough to indicate that the external defects can be grouped visually into four classes (Fig. 3): Skewed Left histogram ('granizo' and 'serpeta'); Normal (undamaged); Bimodal ('molestado') and Skewed Right (purple, wrinkled, and rehús). This categorization is used for further discrimination purposes.

Fig. 2 shows the covariance of histogram intensity levels (256) for the Red channel. Similar figures were created for the other channels (not shown). The graph shows the symmetry of the matrix where both the rows and columns are reflected by gray levels. The intensity bar on the right shows the covariance magnitude as represented by the colours in the plot. Higher covariance magnitude indicates a positive covariance while a lower magnitude indicates a negative covariance between the pairs of variables. Two gray levels (x, y) in each graph were selected corresponding to the highest variance. The resulting variables are (expressed with the initial letter of the colour parameter and the number of the gray level selected): Red: R70,

Table 2
Results of ANOVA for "Manzanilla Sevillana" olives classified into damage type studied

Defect type	N	Olives/kg	g/fruit	Height (mm)	Equatorial diameter (mm)	L^*	a^*	b^*
Purple	68	200	5.0 (1.0)^{a}	19.3 (1.5)^{a}	22.5 (1.9)^{a}	28.1 (2.7)^{d}	8.1 (3.7)^{a}	2.20 (2.4)^{e}
'Rehús'	23	227	4.4 (1.0)^{b}	18.9 (1.6)^{ab}	22.5 (1.7)^{a}	42.2 (7.8)^{c}	6.0 (5.8)^{ab}	17.49 (9.2)^{d}
'Molestado1'	49	250	4.0 (0.7)^{bc}	17.9 (1.3)^{c}	21.4 (1.5)^{b}	53.2 (6.8)^{b}	−0.2 (8.2)^{c}	29.58 (9.2)^{c}
'Serpeta'	11	263	3.8 (0.8)^{bcd}	17.9 (1.4)^{bcd}	20.1 (1.5)^{c}	58.7 (5.4)^{a}	−3.4 (6.6)^{cd}	36.18 (8.7)^{ab}
'Molestado2'	48	278	3.6 (0.8)^{d}	17.0 (1.5)^{d}	20.6 (1.9)^{c}	59.6 (5.2)^{a}	−5.4 (6.5)^{d}	37.31 (7.3)^{ab}
Undamaged	30	278	3.6 (0.7)^{d}	17.0 (1.4)^{d}	20.8 (1.3)^{bc}	59.8 (3.8)^{a}	−8.0 (2.6)^{e}	39.47 (3.9)^{a}
'Granizo'	12	286	3.5 (0.7)^{cd}	17.3 (1.5)^{cd}	20.5 (1.6)^{bc}	56.6 (7.3)^{a}	−2.9 (8.0)^{cd}	34.65 (10.1)^{b}
Wrinkled	19	526	1.9 (0.6)^{e}	13.6 (1.5)^{e}	17.0 (1.8)^{d}	29.7 (2.4)^{d}	3.3 (2.8)^{b}	2.45 (2.3)^{e}
F of Fisher			36.8	39.7	25.1	253.0	38.3	197.1

Mean (standard deviation).
Note. Different letters (a–e) determine the significant differences between group means; same letters in the same columns, there is no statistical significance between the defects ($p > 0.5$).

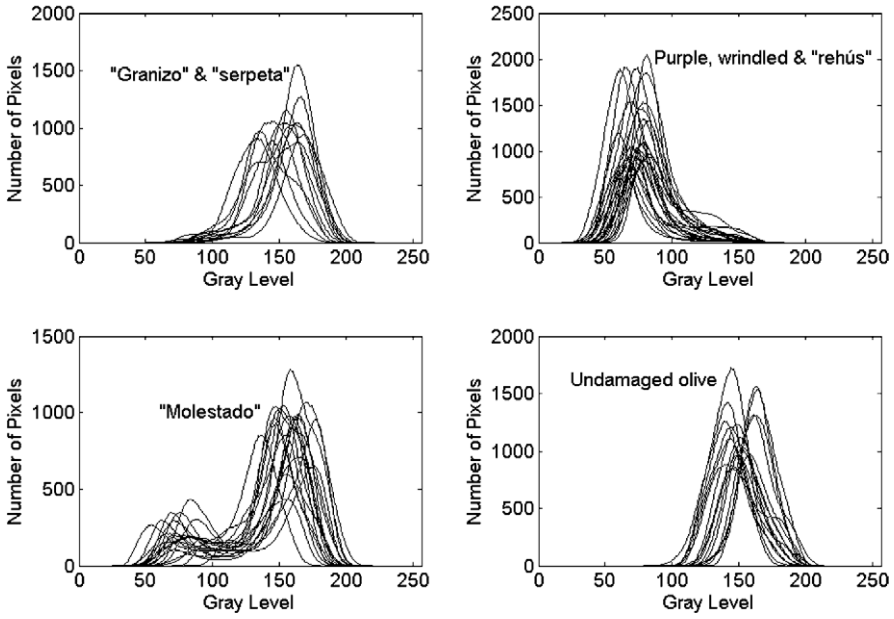

Fig. 3. Example of histograms obtained for all gray levels (256) in the red channel, for the different categories of olives with calibration sets. Each curve corresponds to an olive. The shape of distributions was used to establish groups.

R158; Green: G66, G168; Blue: B71, B110; Hue: H59, H199; Saturation: S30, S102; and Value: V72, V168.

A Fisher LSD test was used to determine the significant differences between group means in an analysis of variance. The purpose of this test was to reduce the number of parameters analyzed, eliminating those variables which were totally interrelated within homogeneous groups.

Other authors (Diaz et al., 2004) have used different grading techniques to sort olives into four classes, getting the best results with neural networks (90%) compared to partial least squares and Mahalanobis distance (70%). However in this work the discriminant analysis technique (DA) was used because it requires a low number of variables to create the functions, a low computational power and it has yielded good results in previous studies (Valero et al., 2004; Hernández-Sánchez et al., 2006).

Consecutive DAs in our work included combinations of colour and morphological features in agreement with several authors, to improve discrimination (Kondo et al., 2000; Leemans and Destain, 2004; Mateos et al., 2005; Unay and Gosselin, 2006). Until now the image analysis of small fruits has been based on colour appearance (Uthaisombut, 1996; Diaz et al., 2004) or just fruit shape. In this work, the blended use of parameters related to defect shapes, fruit colours, and histogram curves was proposed for olives.

In a first discriminant (DA1), the features with the highest discriminatory power between global defects and local defects were included in the resulting linear model:

– One histogram feature: Olive Area
– Several colour features: S30, G66, V72, B110 and B71.

The calibration model was developed with 123 olives: 48 global defects and 75 local defects, with an excellent percentage of correct classification for both types of

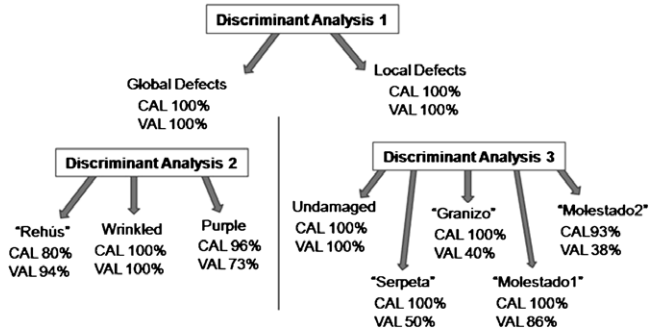

Fig. 4. Classification into different categories of olives for three DAs performed. The percentages refer to fruit classification success rates: Calibration (CAL) and Validation (VAL).

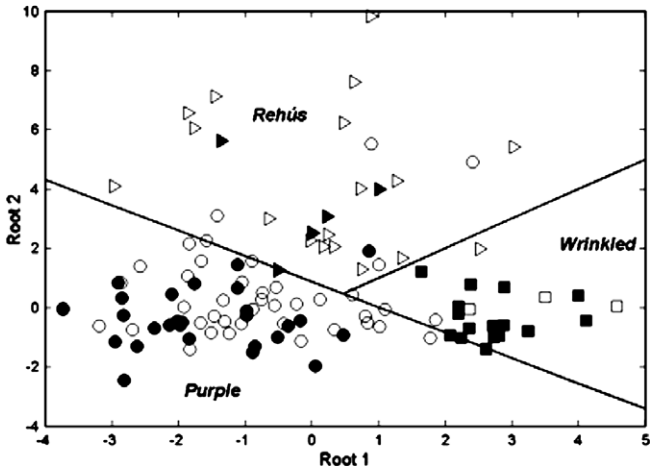

Fig. 5. Plot of canonical discriminant function for the DA2. Calibration (solid symbol) and Validation values (empty symbol) for global defect. □ Wrinkled, ○ Purple and △ *Rehús*.

defects (Fig. 4). The model was validated using the whole set of olives ($N = 137$), also with all fruits correctly classified. Kavdir and Guyer (2004) also worked with histogram features and statistical classifiers, but their results in apples were significantly lower. However, in our work histogram features have enhanced discriminatory power when combined with other parameters in the DA1, as well as in consecutive DAs, as explained in the following paragraphs.

A second discriminant analysis was performed (DA2) trying to distinguish between the global defects. The variables selected for the linear model with higher discriminatory power between global defects were V72, B71, H59 and Olive Area. The addition of a morphological feature (Area) in DA2 was very positive. In accord with the findings of Unay and Gosselin (2006), the addition of a local feature (the intensity of pixels in their case) improved defect segmentation.

The correct classification percentage of global defects (Fig. 4) show that Wrinkled olives are the only well-classified category in this group, probably due to the inclusion of the 'olive area' variable in the model (these fruits are smaller) and their dark colour. The purple olives are sometimes difficult to segregate from 'rehús' when they have several hits on their skin surface. This is why in Fig. 4, the validation score of 'rehús' shows a percentage increase, as some purple olives are included.

Table 3 shows the correlation between the features ultimately selected by the DA2 model. There are higher correlations between V72 and B71 of $r = 0.94$, and also between B71 and Area olive of $r = 0.78$. Despite these correlation levels indicate a certain association between the corresponding variables, they were included in the model because they fulfil the tolerance value set (0.01).

Table 3
Correlations between colour features for the DA2

	V72	B71	H59	Area olive
V72	1			
B71	0.94	1		
H59	0	−0.01	1	
Area olive	0.67	0.78	0.08	1

Fig. 5 represents the olives studied in DA2, on a two canonical function space (root 1 and root 2). As the number of defect categories to classify in the DA2 is three, two root functions are necessary. The three categories are reasonably well segregated by the function lines in the graph.

Table 4 show the χ^2 test results for both successive roots of DA2, in order to analyze which discriminant canonical function gives better discrimination. The table contains the significance test for the two roots ($p < 0.01$), being both significant. More parameters were also computed, as the eigenvalues (meaning the cumulative proportion of the variance explained by each root), as well as the canonical correlation, Wilks' lambda (contribution of each root to the overall performance) and degree of freedom. The first function yields a discrimination rate of 76% between groups while the second function yields one of 90%. Both are highly effective at discrimination. Fig. 5 shows the three groups of defects sorted in this DA2, which are separated reasonably well according to the percentages in Fig. 4.

A third discriminant analysis (DA3) was necessary to classify among the local defects. In this case the variables introduced in the stepwise discriminant analysis belong to colour features and defect features. Finally, the attributes with higher discriminatory power between global defects and local defects were 'Eccentricity', 'EquivDiameter', 'MinorAxisLength', 'Solidity', 'Roundness', 'MajorAxisLength', 'Number of objects' and the olive gray level V30.

The percentages of correctly classified olives are summarized in Fig. 4, both via calibration and validation. The calibration of all categories, with the exception of 'molestado2', have a 100% score. For validation, only undamaged olives kept this value.

Table 4
χ^2 tests with successive roots extracted for the DA2

	Eigen value	Canonical *R*	Wilks' *Λ*	χ^2	Degrees of freedom	*p*-Level
1	4.06	0.90	0.08	107.71	8	0.000
2	1.35	0.76	0.43	37.15	3	0.000

Some fruits presented a combination of various defects, which explains the difficulty of separating them by categories. However, observations of the original classification matrix indicate that the olives included in the 'molestado' groups are almost always well classified. The reason for the weak classification with the rest of the defects could be the poor observation numbers.

For 'serpeta' and 'molestado2', both defects have similar geometry and colour (Fig. 1: olives "b" and "i-left"). Thus, classification scores of 'serpeta' and 'molestado2' reach 100% for calibration (50% validation) and 93% (38% validation). Regarding the 'granizo' group, the low score for validation (40%) may be due to the scarce number of samples in both sets: 7 in calibration and 5 in validation. This is one of the main limitations of the present study, and clearly affects the final result.

In this third discriminant analysis (DA3), there are also higher correlations between some features used (Table 5), 0.93 being the highest correlation coefficient, between 'EquivDiameter' and 'MinorAxisLength', followed by 0.89 between 'Roundness' and 'Number of objects'. As in DA2, the tolerance was set to 0.01.

As in DA2 (Table 4), Table 6 shows the values corresponding to each canonical function extracted in DA3. The numbers of defect categories to classify in the DA2 are five. Thus, four canonical functions are obtained. In this case, all the functions explain the high percentage of discrimination between groups. Fig. 6 shows the best representation of a canonical discriminant analysis of global defects for DA3, in a root 1 to root 2 space.

The undamaged olives are perfectly segregated from the rest of the categories included in DA3; although apparently in the Fig. 6, the "undamaged" samples seem to be fewer than the real number (16 calibration, 14 validation), this is only an optical effect. Almost all values of undamaged olives were represented on the same point (with minimal differences), due to the fact that all those fruits were free of defects (being all fruit suitable to be marketed). Therefore the algorithm used for defect extraction did not detect any faulty area in them, and consequently all features included in DA3 ('Eccentricity', EquivDiameter', 'MinorAxisLength', 'Solidity', 'Roundness', 'MajorAxisLength', and 'Number of objects') had a value equal to cero (*y* axis in Fig. 6, root 2) except the features 'olive gray level V30' (*x* axis, root 1). There were only some olives of the validation set which did not fulfill the last conditions (i.e. small area of purple colour in the skin, detected as a defect by the algorithm), nevertheless the feature classified them as undamaged olives.

According to Fig. 4 and Table 6 the canonical correlation value of root function 1 is 100%. Keeping in mind that several olives presented a mixture of defects, 'molestado1' segregates reasonably well (100% calibration, 86% validation).

Table 6
χ^2 tests with successive roots extracted for the DA3

	Eigenvalue	Canonical *R*	Wilks' *Λ*	χ^2	Degrees of freedom	*p*-Level
1	255.37	1.00	0.00	523.68	32	0.000
2	3.41	0.88	0.11	149.28	21	0.000
3	0.81	0.67	0.48	49.14	12	0.000
4	0.14	0.35	0.87	9.02	5	0.108

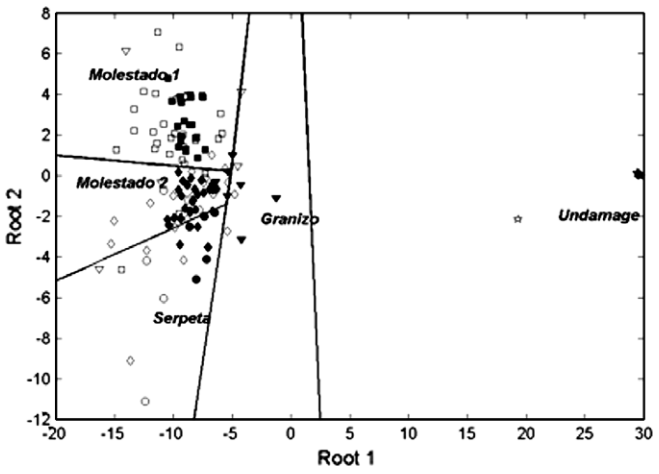

Fig. 6. Plot of canonical discriminant function for the DA 3: Root1 and Root2. Calibration (solid symbol) and Validation values (empty symbol) for global defect. ☆ Undamaged, □ '*Molestado1*', ◇ '*Molestado2*', ○ ' *Serpeta*' and △ '*Granizo*'.

Table 5
Correlations between segmentation features for the DA3

	Eccentricity	Equiv. diameter	MinorAxisLength	Solidity	Roundness	MajorAxisLength	Number of objects
Eccentricity	1						
Equiv. Diameter	0.53	1					
MinorAxisLength	0.46	0.93	1				
Solidity	0.64	0.61	0.40	1			
Roundness	0.49	0.53	0.69	0.08	1		
MajorAxisLength	0.73	0.83	0.84	0.39	0.74	1	
Number of objects	0.44	0.52	0.68	0.04	0.89	0.7	1

4. Conclusions

A procedure is proposed that enables the identification of sound olives as well as a variety of defects based on three discriminant analyses. This methodology for the classification of olives makes use of colour features of the fruit together with several morphological characteristics of external defects, which enhances the final performance.

Finally, olive classification into eight classes was adequately achieved. The characterization of olives by means of image sets aims to provide reliable references which will be used to evaluate the vision devices of on-line classification equipment.

It would be advisable to perform a specific study with a larger number of samples, although is difficult find more samples for some defect types.

Acknowledgements

The research included in this paper has been carried out at the Polytechnic University of Madrid (Spain), with support provided by the Comunidad Autónoma de Madrid as part of the OPTICAM (07G//0014/2003 1) research Project and the common research action "TAGRALIA." Thanks are also due to MULTISCAN S.L. Company which designed the ESCA200 vision equipment, and collaborates with LPF-TAG on a common project, as well as to several olive factories that supplied the plant material.

The authors would like to acknowledge the "Commission Internationale du Génie Rural" (CIGR) and to commend their work through the "Armand Blanca Prize 2006".

References

Alvarado, M., Civantos, M., Duran, J.M., 2001. Plagas/Pests. In: Barranco, D., Fernández-Escobar, R., Rallo, L. (Eds.), El cultivo del olivo/the olive crop. Ed. Grupo Mundi-Prensa, Madrid, Spain, pp. 435–493.

Barreiro, P., Moya, A., Ruiz-Altisent, M., Agulheiro, A. C., García-Ramos, F. J., Homer, I., Moreda, G., 2003. Separación de aceitunas sobremaduras y alambradas en línea mediante rebote (Segregating over-ripe and fish-eyed ('alambradas') olives on-line by means of rebound). In: II Congreso Nacional de Agroingenieria, Córdoba, Spain.

Barreiro, P., Zheng, C., Sun, D.W., Hernández-Sánchez, N., Pérez-Sánchez, J.M., Ruiz-Cabello, J., 2008. Non-destructive seed detection in mandarins: Comparison of automatic threshold methods in FLASH and COMSPIRA MRIs. Postharvest Biology and Technology 47 (2), 189–198.

Bennedsen, B.S., Peterson, D.L., 2005. Performance of a system for apple surface defect identification in near-infrared images. Biosystem Engineering 90 (4), 419–431.

Blasco, J., Aleixos, N., Moltó, E., 2007. Computer vision detection of peel defects in citrus by means of a region oriented segmentation algorithm. Journal of Food Engineering 81, 535–543.

Brosnan, T., Sun, D.W., 2004. Improving quality inspection of food products by computer vision-a review. Journal of Food Engineering 61, 3–16.

Cheng, X., Tao, Y., Chen, Y.R., Luo, Y., 2003. NIR/MIR dual-sensor machine vision system for online apple stem-end/calyx recognition. Transactions of ASAE 46 (2), 551–558.

Cordero, S., Lleó, L., Barreiro, P., Ruiz-Altisent, M. 2006. Peach multispectral images related to firmness and maturity. In: World Congress: Agricultural Engineering for a Better World. Bonn, Germany.

Diaz, R., Faus, G., Blasco, M., Blasco, J., Moltó, E., 2000. The application of a fast algorithm for the classification of olives by machine vision. Food Research International 33, 305–309.

Diaz, R., Gil, L., Serrano, C., Blasco, M., Moltó, E., Blasco, J., 2004. Comparison of three algorithms in the classification of table olives by means of computer vision. Journal of Food Engineering 61 (I), 101–107.

Du, C.J., Sun, D.W., 2004. Recent developments in the applications of image processing techniques for food quality evaluation. Trends in Food Science and Technology 15, 230–249.

FAO, 1987. Norma del CODEX para las aceitunas de mesa, CODEX alimentarious. CODEX STAN 66-1981, (Rev. 1-1987), 1–19.

Hernández-Sánchez, N., Barreiro, P., Ruiz-Cabello, J., 2006. On-line identification of seeds in mandarins with magnetic resonance imaging. Biosystem Engineering 95 (4), 529–536.

Kavdir, I., Guyer, D.E., 2004. Comparison of artificial neural networks and statistical classifiers in apple sorting using textural features. Biosystem Engineering 89 (3), 331–344.

Kleynen, O., Leemans, V., Destain, M.F., 2005. Development of a multi-spectral vision system for the detection of defects on apples. Journal of Food Engineering 69, 41–49.

Kondo, N., Ahmad, U., Monta, M., Murasc, H., 2000. Machine vision based quality evaluation of Iyokan orange fruit using neural networks. Computers and Electronics in Agriculture 29, 135–147.

Leemans, V., Destain, M.F., 2004. A real-time grading method of apples based on features extracted from defects. Journal Food Engineering 61, 83–89.

Mateos, F., Patiño, A., Pérez, J., Tombs, J., Aguirre, M.A., 2005. Implementación en FPGA de un sistema de procesamiento en tiempo real para la detección de defectos en frutos/the implementation in FPGA of a real-time processing system for the detection of defects on fruit. In: V Jornadas de Computación Reconfigurable y Aplicaciones (Jcra 2005, pp. 151–157). Granada, Spain.

Mathworks, 2005. Image Processing Toolbox User's Guide (For Use with MATLAB). <http://www.mathworks.com/>.

Mehl, P.M., Chen, Y.R., Kim, M.S., Chan, D.E., 2004. Development of hyperspectral imaging technique for the detection of apple surface defects and contaminations. Journal Food Engineering 61, 67–81.

Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE Transactions Systems, Man, and Cybernetics smc-9 (1), 62–66.

Riquelme, M. T., Ruiz-Altisent, M., Valero, C. 2004. Informe sobre la caracterización de aceitunas "var. Manzanilla Sevillana"/Report about characterization of "Manzanilla sevillana" olives (Confidential report).

Rosenberger, C., Emile, B., Laurent, H., 2004. Calibration and quality control of cherries by artificial vision. Journal of Electronical Imaging 13 (3), 539–546.

StatSoft, Inc., 2007. Electronic Statistics Textbook. Tulsa, OK, USA: StatSoft. <http://www.statsoft.com/textbook/stathome.html>.

Trapero, A., Blanco, M.A., 2001. Enfermedades/diseases. In: Barranco, D., Fernández-Escobar, R., Rallo, L. (Eds.), El cultivo del olivo/the olive crop. Ed. Grupo Mundi-Prensa, Madrid, Spain, pp. 497–550.

Unay, D., Gosselin, B., 2006. Automatic defect segmentation of 'Jonagold' apples on multi-spectral images: A comparative study. Postharvest Biology and Technology 42, 271–279.

USDA, 1967. United States Standards for Grades of Green Olives. United States Department of Agriculture Washington, Federal Register August 9, 1967 (32 F.R. 11467).

Uthaisombut, P. 1996. Detecting defects in cherries using machine vision. M.S. Thesis, Computer Science, Michigan State University.

Valero, C., Ruiz-Altisent, M., Cubeddu, R., Pifferi, A., Taroni, P., Torricelli, A., Valentini, G., Johnson, D., Dover, C., 2004. Selection models for the internal quality of fruit, based on time domain laser reflectance spectroscopy. Biosystem Engineering 88 (3), 313–323.