



On-line Fruit Grading according to their External Quality using Machine Vision

V. Leemans^a; H. Magein^b; M.-F. Destain^a

^aLaboratoire de mécanique agricole, Gembloux Agricultural University, Passage des Déportés, 2, B-5030 Gembloux, Belgium;
e-mail of corresponding author: leemans.v@fsagx.ac.be

^bDepartment of Biotechnology, Agricultural Research Centre of Gembloux, Chaussée de Charleroi, 234, B-5030 Gembloux, Belgium;
e-mail: magein@cragx.fgov.be

(Received 24 September 2001; accepted in revised form 13 August 2002)

This paper presents apple grading into four classes according to European standards. Two varieties were tested: Golden Delicious and Jonagold. The image database included more than a 1000 images of fruits (528 Golden Delicious, 642 Jonagold) belonging to the three acceptable categories—Extra, I and II—and the reject (each class represents, respectively, about 60, 10 and 20% of the sample size).

The image grading was achieved in six steps: image acquisition; ground colour classification; defect segmentation; calyx and stem recognition; defects characterisation and finally the fruit classification into quality classes.

The proposed method for apple external quality grading showed correct classification rates of 78 and 72%, for Golden Delicious and Jonagold apples, respectively. Taking into account that the healthy fruit were far better graded and considering that this class was under represented in the sample compared with the fruit population, the results of the proposed method (an error rate which drops to 5 and 10%, respectively) are compatible with the requirements of European standards. © 2002 Silsoe Research Institute. Published by Elsevier Science Ltd. All rights reserved

1. Introduction

Apples are graded into three categories or rejected, according to European standards (Anonymous, 1989), on the basis of their external quality. Besides colour and shape specifications, this involves the presence of defects, their origin and their number or size. The category 'Extra' should have no defects, category 'I' allows small skin defects while category 'II' permits fruits with more serious defects. Fruits, which do not reach the minimum requirements, are rejected and are traditionally used in processing. The automation of the grading could be achieved by colour machine vision. This would require at least three steps: image acquisition, its segmentation (to locate the fruit, the defects, the background colour *vs* the blush) and the classification of the fruit.

2. Literature review

Several methods focused on fruit classification are described in the literature. Early studies on fruit external quality grading used global parameters. Miller (1995) took into account the mean fruit colour and a measure

of the dispersion (normalised mean squared differences) of the colour, plus a shape parameter to grade citrus fruits according to their external quality. The author compared three different classifiers and had the best results with Bayesian–Gaussian techniques, with between 69 and 86% of the fruit correctly graded into two classes (accepted or rejected). Nakano (1997) used neural networks (two-layer perceptron, five hidden neurons) to sort San-Fuji apples into five colour and quality classes. In a first step, the pixels were classified according to their colour, their position and the mean colour of the fruit. In a second step, the fruit were graded using 11 parameters (fruit mean colour, colour variability, presence of 'defect pixels' and the 'ratio of normal red colour'). The correct classification rate varied from 33 to 95%, according to the class (the global classification rate was about 70%). Guedalia (1997) compared two methods to summarise the data resulting from the segmentation and then graded the apples into four classes (on a large set of 1100 fruits). The author used a supervised method requiring the classification of the objects (blobs resulting from a segmentation, *i.e.* the calyx, the stem or a defects) and

then the grading of the fruits. The second method, unsupervised, computed the principal components from the set of parameters characterising the objects and graded the fruits directly on this set of components. The correct classification rate reached 48% for a treatment by fuzzy logic, and 67% for a classification by neural network. Picus and Peleg (1998) presented a dynamic dates grading method. The key point was that the fruits were graded using a reference population which was not constant, but which was a subset (a cluster) from the whole population. During the grading process, the characteristics of the last n fruits were included in a 'first in, first out heap' which was used to determine the correct sub-set. The error rate dropped from between 65 and 33% for a fixed population to 26% for the dynamic grading. Leemans *et al.* (1998c) used a supervised method to grade the blobs resulting from the image segmentation. Large (above 11 mm²) and small blobs were treated differently. In a first step, for the former the over-segmentation was separated from the other blobs. In the second step, five classes were considered, following European Union standards (main defects, slight defect acceptable in category I or II according to their size, scab, russet and stem end). Twelve parameters measuring the shape, the colour and the texture were used in linear or quadratic discriminant analysis as well as in neural networks. Quadratic discriminant analysis gave the best results, with between 43 and 100% (function of the class) of the blobs correctly classified. For the small blobs, only five parameters and three classes were considered. Linear discriminant analysis gave the best results with a correct classification rate of 77% of the sample. This method was supervised, thus requiring an operator to grade several thousand blobs, which slows down the process considerably.

Across these studies, it appears that the fruit grading methods could be ordered from global ones, computing descriptors directly on the image, up to methods requiring several steps: defect detection (image segmentation), condensation of the information and fruit classification. The latter seems more adapted to an accurate grading of apples, because of the presence of a calyx and a stem end, and of the colour variation present on some fruits, especially the bi-colour ones. The method presented in this paper belongs to the second group of methods, a hierarchical approach taking into account the whole information available.

3. Materials and methods

Two varieties were studied, the mono-colour Golden Delicious, and the bi-colour Jonagold. The data characterising the fruit are given in Table 1.

Table 1
Characteristics of the two varieties data sets

<i>Data type</i>	<i>Golden Delicious</i>	<i>Jonagold</i>
Number of images	528	642
Extra fruits, %	69	57
Category I fruits, %	13	27
Category II fruits, %	6	5
Reject fruits, %	13	11

Healthy fruits are underrepresented in the sample size compared with the population, so as to not overload the image database, while representing each defect or class with enough images (especially category 'II' which represents only several per cent of the whole population). The defects encountered over 3 years (1997–1999) were: bruises, wounds (impacts, birds), scab, russet, fungal attack, bitter pit, scar tissue, frost damage, insect attack.

The fruit grading was achieved in six steps: image acquisition; colour classification; defect segmentation; calyx and stem end recognition; defects characterisation and finally the fruit classification into quality classes.

3.1. Image acquisition

The image acquisition system is described by Leemans (1999) and is made of a lighting tunnel, two cameras and two frame grabbers. The light was emitted by two lighting tubes placed below the level of the fruit and was reflected by the inner surface of the tunnel. This tunnel was painted flat white and gave diffused and homogeneous lighting. The fruits were placed on rollers able to move through the tunnel with a chosen horizontal rotational speed (quite classical for fruit grading machines).

Two cameras (colour, three charge-coupled devices, Sony XC003P) were placed in a plane perpendicular to the tunnel axis and were inclined at 45°.

The basic image treatments were made using Easylib (Euresys S.A.) libraries, incorporated into a C++ program, which was also used to develop the complementary process. The statistics and the discriminant analysis were made using Minitab (Minitab Inc.), while the neural networks were developed with Matlab (Mathworks Inc.).

3.2. Colour grading

Leemans *et al.* (1997, 1998b) have described the ground colour classification. For Jonagold apples (bi-colour fruit), the ground colour and the blush had to be

separated first. This was best achieved using a simple neural network with no hidden layer, using the three luminances (*red*, *green* and *blue*) of the considered pixel as input. The ground colour classification—grading of a fruit into four classes from far green to yellow—was achieved using Fisher's linear discriminant analysis for both Golden Delicious and Jonagold. The 'predictor' was the first canonical variate, computed on the mean value of the three luminances estimated on the ground colour area for Jonagold or on the whole surface of the fruit for Golden Delicious.

3.3. Calyx and stem end recognition

The calyx and stem ends, which appear on an image as defects, were detected using a correlation pattern recognition technique (provided in the Easylib libraries, described by Leemans, 1999). Two reference images, one for the calyx end and one for the stem end, were built averaging five images for each end. The algorithm returned the gravity centre of the detected object and a correlation score. This parameter could vary between 0 when there is no relation between the target and the image being analysed and 1 when both match perfectly.

3.4. Image segmentation

Leemans *et al.* (1998a, 1999) presented the image segmentation techniques used for the defect detection, respectively, for Golden Delicious apples and Jonagold. The first one used a Gaussian model of the fruit colour, measuring the Mahalanobis distance separating the mean colour of the fruit and of each pixel. For the Jonagold apples, having a multimodal colour frequency distribution, the defect location was based on a non-parametric model of the fruit colour and on Bayes's theorem. In both cases, the development of an algorithm, taking into account local information, enhanced the segmentation precision. From the image segmentation resulted a three (Golden Delicious) or four (Jonagold) grey-level image (examples in *Fig. 1*). The black level was attributed to the background, two grey levels were used for the fruit ground colour and blush. The pixels presented in white included the defects, the calyx and stem ends and some over-segmentation. These areas are called blobs within the framework of this paper.

3.5. Defects characterisation

The amount of blobs varied considerably, from zero for a fruit with no defects up to a 100. The data

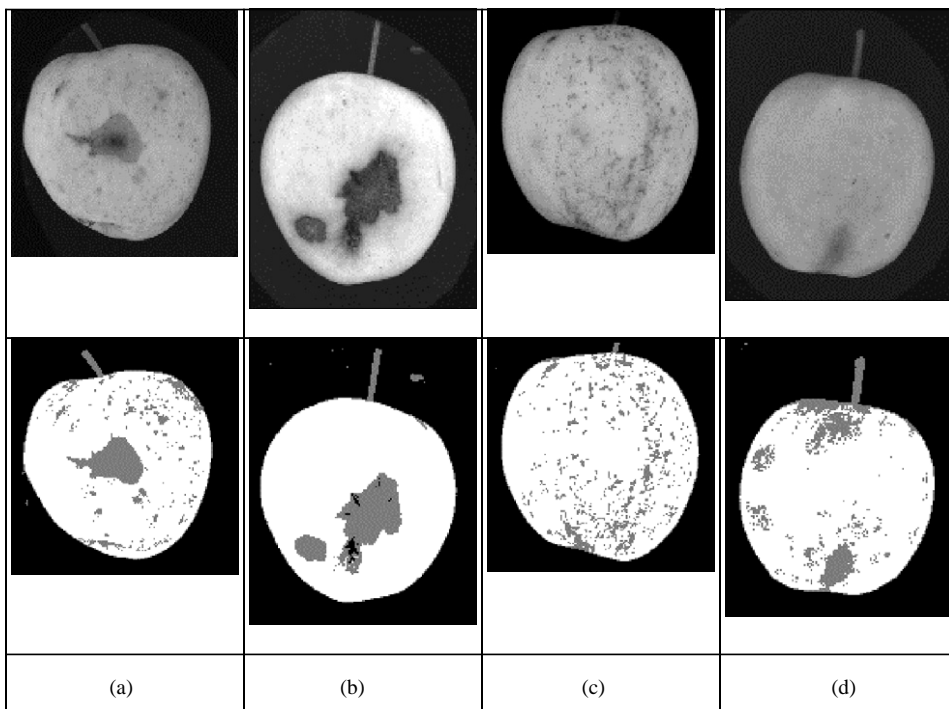


Fig. 1. Samples of Golden Delicious apples with different defects segmented using a Gaussian model of the fruit colour: (a) typical defect; (b) well-contrasted defect; (c) diffuse defect and (d) bruise

computed to describe the blobs, forming a dynamic table, had to be summarised into a static table before being introduced into a classifier. The main difference between this method and the one proposed by Leemans *et al.* (1998a) is that the blobs are grouped by 'k-means' clustering instead of being classified by an expert. This procedure classifies observations into groups when the groups are initially not known. The number of clusters n_c must be fixed. In a first step, the centre of the clusters (the mean values) are given by the parameters of the n_c first objects. Then all the objects are classified in the nearest cluster (according to the Euclidean distance) and the mean values of the parameters are re-computed. This step is repeated up to convergence. The clustering procedure is established using the data of the whole set of fruit. Next, in order to grade each fruit, the information included in the clusters needed to be condensed. This was achieved using the classification probability of the blobs into the clusters. The fruits were then finally graded by linear discriminant analysis or by neural networks.

Several groups of parameters were considered suitable to characterise a defect. These were

(1) *Geometrical parameters (shape, size and position).*

The area A , the fourth root of the area $A^{1/4}$, the perimeter p , the square root of the perimeter $p^{1/2}$, the major inertia moment I_{maj} , the ratio of the inertia moments R_I , the mean diameter d_m , the circularity C ($C = P^2/(4A\pi)$), the length along the main inertia axis L , the width along the minor inertia axis W , the distance from the gravity centre of the fruit to the gravity centre of the object l .

(2) *Colour parameters.* The mean value for the red channel r_m , the mean value for the green channel g_m , the mean value for the blue channel b_m , a colour index representing the background colour G , the Euclidean distance between the fruit background colour and the defect mean colour $l_{G,D}$.

(3) *Texture parameters.* The standard deviation for the red channel r_{sd} , the standard deviation for the green channel g_{sd} , the standard deviation for the blue channel b_{sd} , the mean value for the gradient computed on the red channel r_{grad} , the standard deviation for the gradient computed on the red channel $r_{grad,sd}$.

(4) *Parameters related to the calyx and stem ends.* The score for the stem end detection S_{ped} , the score for the calyx end detection S_{cal} , the distance between the considered blob and the detected stem end l_{ped} , the distance between the considered blob and the detected calyx end l_{cal} .

The selection of the parameters was made in a first instance examining the correlation matrix and in the second using a stepwise selection.

Afterwards, clusters were created in the parameter hyperspace. A cluster was defined by its shape, colour, texture or the distance from the calyx or the stem end, rather than by the origin of the defect. The parameters selected at the previous point were used and all the blobs resulting from the segmentation of the training set were considered simultaneously. The data were standardised (divided by their standard deviation) in order to give them all the same weight. The best number of clusters was unknown. In general, increasing this number could enhance the accuracy of the fruit grading, but this is limited by the sampling size. The number of clusters considered was between 6 and 20.

Two parameters were computed to characterise each image. The first was the sum S_{pk} of the *a posteriori* probability P_{ki} for cluster k :

$$S_{Pk} = \sum_{i=0}^{n_c-1} P_{ki}$$

The second one was the variance of the *a posteriori* probability P_{ki} for cluster k :

$$V_{Pk} = \frac{1}{n_c} \sum_{i=0}^{n_c-1} (P_{ki} - \bar{P}_k)^2$$

These two indices were computed for each of the n_c clusters, giving $2n_c$ indices.

3.6. Fruit grading

Two grading methods were compared:

- (1) quadratic discriminant analysis preceded by the computation of the n_{pc} first principal components (used for data reduction);
- (2) neural network, with a multi-layer perceptron with one hidden layer.

In both cases, the images had to be classified into four classes: the three categories Extra, I, II and the reject. The grading error was estimated on the validation set.

A final test showed how the fruit could be graded into two classes (accepted *versus* rejected). As the *a priori* probability of the different classes in the samples was not representative of what could be observed at the grading station, the influence of the ratio of healthy fruits was also studied.

4. Results and discussion

From the starting set of parameters able to characterise the blobs, 18 were finally selected. These were, the fourth root of the area $A^{1/4}$, the square root of the perimeter $p^{1/4}$, the major inertia moment I_{maj} , the ratio

of the inertia moments R_I , the distance from the gravity centre of the fruit to the gravity centre of the object l , the mean value for the red channel r_m , the mean value for the green channel g_m , the mean value for the blue channel b_m , a colour index representing the background colour G , the Euclidean distance between the fruit background colour and the defect mean colour $l_{G,D}$, the standard deviation for the red channel r_{sd} , the standard deviation for the green channel g_{sd} , the standard deviation for the blue channel b_{sd} , the mean value for the gradient computed on the red channel r_{grad} , the standard deviation for the gradient computed on the red channel $r_{grad,sd}$, the score for the stem end detection S_{ped} , the score for the calyx end detection S_{cal} , the distance between the considered blob and the detected stem end l_{ped} and the distance between the considered blob and the detected calyx end l_{cal} . The correlation matrix between those data showed the linkage between the three colour parameters (*sensus stricto*, i.e. r_m, g_m, b_m —the correlation coefficient r is about 0.8), those related to the size ($A^{1/4}, p^{1/2}, I_{maj}$ — r about 0.9) or those pertinent to the texture ($r_{sd}, g_{sd}, b_{sd}, r_{grad}, r_{grad,sd}$ — r from 0.5 to 0.8). However, removing one or several of these variables caused a drop in the fruit classification rate. The number of clusters n_c giving the best classification rate was found

to be 12 for Golden Delicious and 16 for Jonagold. Fifteen principal components were used for the fruit grading (n_{pc}).

Figure 2 shows scatter diagrams for three parameters for the variety Jonagold. The diagram indicates that the link between the different parameters was more complex than a linear relation but revealed several poles. The effect of the different families of parameters can be clearly observed by noticing the elongation of the scatter diagram along the three axes to:

- the effect of the size ($A^{1/4}$);
- the effect of the luminance (r_m) and
- the effect of the texture (r_{sd}).

Some clusters (2, 4, 5 and 6) can be easily identified by a high value in one or two parameter families (e.g. cluster 4 included bright and small blobs while cluster 5 included big and bright blobs). Most clusters (7, 9, 10 and 16) had, however, average or small values. This can be explained by the fact that most of the variables presented an I-shaped frequency distribution (not shown). The classification of a blob into these clusters showed *a posteriori* probability P_{ki} often smaller. For these kind of clusters, other parameters like l, R_I , the detection scores, etc. made the difference between clusters. For example (not shown), elongated objects

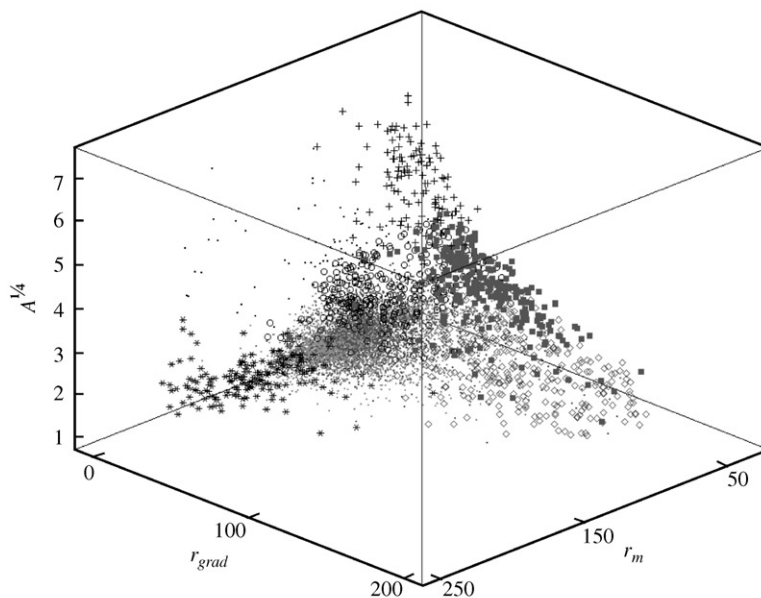


Fig. 2. Scatter diagram showing the clusters of objects segmented on the fruit images for three parameters: the fourth root of area ($A^{1/4}$); the mean value of the red channel (r_m) and the mean value of the gradient computed on the red channel (r_{grad}). Due to the large amount of objects represented on the diagram and to the multiples parameters used to make the clustering, only several clusters can be observed: black circles, cluster 1; black plus, cluster 2; black stars, cluster 4; black dots, cluster 5; light grey diamond, cluster 6; dark grey filled squares, cluster 13. The objects belonging to all the other clusters are represented using light or dark grey dots

(high value for R_I) were found in clusters 8 (dark) and 12 (bright).

Figure 3 gives an illustration of the distribution of the defects from different origins for three clusters. Usually there was no strong relationship between one cluster and a particular origin. For example, most defects could be found in cluster 2. However, this was not important for the grading of the fruit. Cluster 2 included defects with an area bigger than 250 mm^2 . The standard requires that a fruit with such a defect area should always be rejected, no matter the origin of the defect, and this was what the classifier did. On the other hand for other clusters, the influence on the fruit classification depended on the number of objects. Few objects belonging to cluster 8, 12 or 15 had any influence on the grading of the fruit. An increasing number of them had a larger impact on the fruit classification. Cluster 10 was never encountered on healthy fruit, but neither was it a defect. This implies that the colour of the tissue around a defect was influenced by the presence of the defect. Other clusters were typically over-segmentation (3, 4, 7 and 14) and were found to have no influence on fruit classification.

The results of the fruit classification by quadratic discriminant analysis are given in Table 2 (Golden) and

Table 3 (Jonagold). The global correct classification rates were 78 and 72%, respectively, for Golden and Jonagold. Neural networks gave similar results (respectively, 79 and 70%).

The fruit in category 'Extra' were far better graded than those belonging to other classes. This can be easily explained, considering that fruit in these category presented no blob after segmentation (S_{pk} and V_{pk} were thus null) or blobs belonging to over-segmentation.

The analysis of the classification errors for the Jonagold apples showed four origins:

- (1) Rosy apple aphid damage (*Dysaphis plantaginea*—Fig. 4) was correctly graded in 13% of the cases, and for 30% of them, there was an error of one class. This damage creates unevenness on the fruit surface rather than a patch and there was *a priori* nothing to segment. The proposed technique would not be used for this kind of defect. The segmentation results often showed on these fruits a lot of small blobs, but these were normally 'considered' as over-segmentation.
- (2) About 60% of the fruit having a recent bruise were poorly graded. A first origin of the errors was the difficulty to segment those defects, linked to the

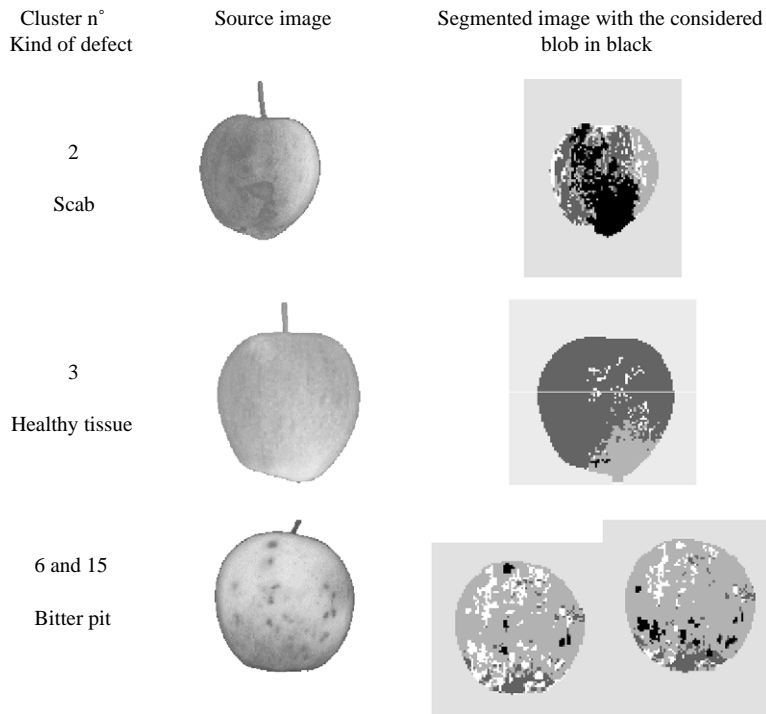


Fig. 3. Examples of typical cluster members; left is the green channel of the source image, right is the segmented image; the background is represented in light grey, the ground colour of the fruit in medium grey, the blush in dark grey and the defects in white except those belonging to the considered cluster represented in black

proximity between the fruit and the defect colours, as shown in Fig. 4. It should be noticed that the fruit with old bruises (present on the fruit before coming into the laboratory) were far better graded than those with a bruise produced in the laboratory and observed 2, 24 or 48 h after an impact.

- (3) Images showing a calyx or a stem end were rejected. These elements were found in two clusters, 13 and 11, the latter for the russet around the stem or calyx end. However, cluster 13 also included some defects having the same appearance.

Table 2
Confusion matrix for the grading of the Golden Delicious apples into four classes

True groups	Image grading class			
	Extra	I	II	Reject
Extra	0.96	0.63	0.00	0.11
Category I	0.03	0.33	0.27	0.09
Category II	0.01	0.03	0.57	0.32
Reject	0.00	0.01	0.17	0.48

Table 3
Confusion matrix for the grading of the Jonagold apples in to four classes

True groups	Image grading class			
	Extra	I	II	Reject
Extra	0.89	0.66	0.23	0.05
Category I	0.04	0.17	0.15	0.04
Category II	0.02	0.04	0.26	0.09
Reject	0.05	0.13	0.35	0.81

- (4) Other errors represented 15% of the whole errors, which was about 4% of the fruit, were considered as residual errors. These were always acceptable in the tolerances premised by the standards.

For Golden Delicious, the errors came mainly from the bruises or were residual errors. Images presenting a calyx or a stem end were correctly graded. When two classes were taken into consideration (fruit accepted or rejected), the error rate reached 5% for Golden Delicious and 8% for Jonagold. The simulation modifying the *a priori* probabilities showed that if the ratio of healthy fruit went up to 90%, the error rate would drop from 22 and 28% (for Golden Delicious and Jonagold, respectively, and for four classes) to 5 and 10%.

This method used an unsupervised grading of the blobs, contrary to the one proposed earlier (Leemans *et al.*, 1998c), and thus eliminated the ponderous task of grading them by an operator (requiring several days). The consequence is an acceleration of the procedure making it possible concentrate on parameter fitting (segmentation method and parameters, selection of the parameters characterising the blobs, amount of clusters, number of principal components, etc.). Furthermore, the enlargement of the database over time becomes possible.

The results showed that errors came mainly from bruises. This was especially the case for Jonagold apples when the bruises were located in the blush area. To improve the performance of a machine vision system with regard to bruise detection, a better segmentation of these defects prior to the classification should be performed. Use of a more appropriate optical system,

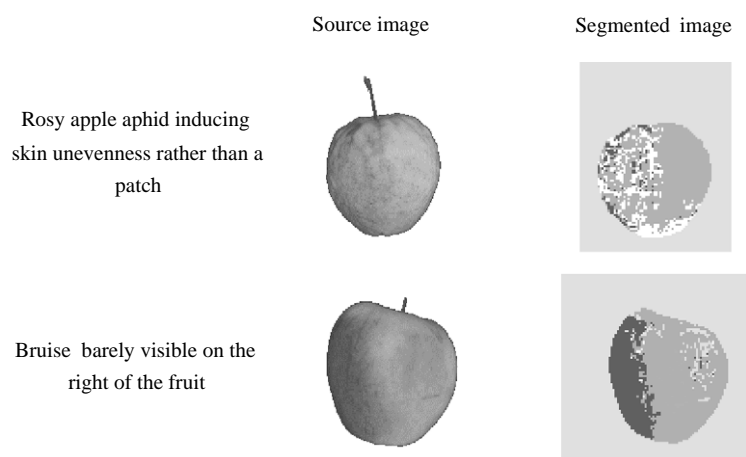


Fig. 4. Examples of images of fruits affected by defects particularly difficult to segment and thus to grade; left is the green channel of the source image, right is the segmented image; the background is represented in light grey, the ground colour of the fruit in medium grey, the blush in dark grey and the defects in white

for example NIR cameras, is suggested. For Jonagolds, it was impossible to detect a special defect which creates unevenness of the fruit surface ('rosy apple aphid damage'). Solving this problem probably implies the use of three-dimensional imaging systems.

Taking into account the performance of the system, its implementation on an industrial machine is possible, even if the problems related to the image acquisition during the fruit's movement have to be solved.

5. Conclusions

This paper has presented a method of grading Golden Delicious and Jonagold apples on the basis of their external defects, these being measured by machine vision. The defects encountered over a 3 year period were represented as bruises, wounds, scab, russet, fungal attack, bitter pit, scar tissue, frost damage and insect damage. The classes corresponding to European standards were considered as Extra, I, II and reject. For Golden Delicious, the correct classification rates reach 78%, by using a database including around 70% of the images category 'Extra'. For Jonagold, this classification rate is 72% with almost 60% of the images category 'Extra'. A simulation shows that, if the ratio of healthy fruits went up to 90%, the correct classification rates would reach 95% and 90% for Golden Delicious and Jonagold apples, respectively.

The defect characterisation was based on parameters describing their geometry, colour, texture and related to calyx and stem ends. The chosen classification method based on the constitution of automatic clusters was found to be efficient. Indeed, it was accurate, as indicated by the classification obtained rates. When compared to methods based on individual defect recognition, it ensures an appreciable saving in time.

Acknowledgements

This research was funded by the Belgian Federal Ministry of Agriculture (project D1/2-5819A Section 1).

References

- Anonymous** (1989). Commission Regulation (EEC) No 920/89. Laying down quality standards for carrots, citrus fruit and desert apples and pears and amending Commission Regulation No 58. Official Journal L097, 11/041 1989 p. 0019-0039
- Guedalia I D** (1997). Visual based quality sorting of fruit. Sensors for Nondestructive Testing—Measuring the Quality of Fresh Fruits and Vegetables. Proceedings from the 'Sensors for Nondestructive Testing International Conference and Tour', Orlando, FL, USA, 18–21 February 1997, North Regional Agricultural Engineering Service, pp 232–238
- Leemans V; Magein H; Destain M-F** (1997). The quality of 'Golden Delicious' apples by colour computer vision. Preprints from the 'Mathematical and Control Applications in Agriculture and Horticulture', International Federation of Automatic Control (IFAC) Workshop, 28 septembre–2 octobre 1997, Hannover, Germany, pp 175–179
- Leemans V; Magein H; Destain M-F** (1998a). Defects segmentation on 'Golden Delicious' apples by using colour machine vision. *Computers and Electronics in Agriculture*, **20**, 117–130
- Leemans V; Bieuvelet F; Destain M-F** (1998b). Application of neural networks in quality control of 'Jonagold' apples. Proceedings of 'Sensoral 98', International Workshop on Sensing Quality of Agricultural Products, 23–27 February 1998, Montpellier, France
- Leemans V; Magein H; Destain M-F** (1998c). Quality fruit grading by colour machine vision: defect recognition. *Acta Horticulturae* 517, Proceedings from 'XXV International Horticultural Congress', 2–7 Aout 1998, Bruxelles, Belgium, pp 405–412
- Leemans V; Magein H; Destain M-F** (1999). Defect segmentation on 'Jonagold' apples using colour vision and Bayesian method. *Computers and Electronics in Agriculture*, **23**, 43–53
- Leemans V** (1999). Contribution to fruit grading by machine vision—application to on-line sorting of Golden Delicious and Jonagold apples. PhD Thesis, Gembloux, Faculté Universitaire des Sciences agronomiques, Belgium, 216 pp, 42 tables, 83 figures
- Miller W M** (1995). Optical defect analysis of Florida citrus. *Applied Engineering in Agriculture*, ASAE, **11**(6), 855–860
- Nakano Kazuhiro** (1997). Application of neural networks to the color grading of apples. *Computers and Electronics in Agriculture*, **18**, 105–116
- Picus M; Peleg K** (1998). Adaptative simulation of human assessment of fruit quality. Proceedings of 'Sensoral 98', International Workshop on Sensing Quality of Agricultural Products, 23–27 February 1998, Montpellier, France