A QUALITY GRADING APPROACH FOR 'JONAGOLD' APPLES

[1]*Devrim Unay, Bernard Gosselin*

[1]unay@tcts.fpms.ac.be
TCTS Lab., Faculté Polytechnique de Mons
Multitel Building, Ave. N. Copernic 1, Parc Initialis, B7000, Mons, Belgium

ABSTRACT

In this paper we introduce a quality grading system with perceptron networks for 'Jonagold' apples. The system is composed of pre-processing, classification, post-processing and decision-taking steps. It is tested by a database of 819 apple images and the resulting images are compared with a reference database which is manually segmented. Performances of 2, 3-class classifiers are compared where 3-class system performed better. An absolute error value is introduced to interpret the classification results at pixel level and the advantage of post-processing is shown.

1. INTRODUCTION

Quality classification and defect segmentation of apples is a hard task due to multi-colored skin of some kinds, stem-calyx parts or concavity of apples leading to misclassifications, and defects highly varying in types, sizes.
In computer vision based apple sorting, Leemans et al.[1]-[2] applied Gaussian and Bayesian-based pixel classification methods on 'Golden Delicious' and 'Jonagold' apples. He also introduced a hierarchical grading method and used k-means clustering for a real-time grading system by which he reached 73% correct classification[3]. Nakano[4] used neural networks for color grading of 'San Fuji' apples from color images. Rennick et al.[5] used a controlled acquisition system and different classifiers for 'Granny Smith' apples. Miller et al.[6] used spectral reflectance properties of apples, whereas Yang and Marchant[7] used snakes and a flooding algorithm to detect blemishes. Wen and Tao[8] built a rule-based, near-infrared, automated vegetable sorting system and reached rates of more than 80%, but misclassified stem and calyx regions as defected. Pla et al.[9] built automated apple sorting system with weight sensors, infra-red, color, and ultra-violet images. In our previous works[10]-[11], we used local and global approaches with principal components analysis[12] to classify apples by perceptron networks from color images, however in the former we lacked a proper database and in the latter the approach was not able to give information on defect size.

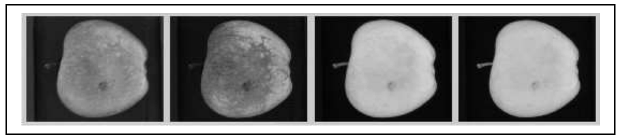

Figure 1: Color-band images of an apple. Left to right: blue, green, red, and infra-red bands.

Our grading system works as follows: A test image is introduced. Pre-processing is applied, samples belonging to apple region are found and their features are extracted. Then these features are introduced to the trained perceptron, which gives a result for each sample. Post-processing is applied on these results to remove very small defect regions. Finally, defect-decision ratio is calculated and the test image is classified as 'I' (accepted) or 'Rejected'.

2. MATERIALS AND METHODS

2.1. Image Database

The database consists of 280 'healthy', 246 'defect', 148 'stem', and 145 'calyx' images of 'Jonagold' apples acquired by a ccd-camera and four filters from one-view in a diffusely illuminated environment (Figure 1) [13]. Some 'healthy' and 'defect' images also contain stem-calyx regions, however 'stem' and 'calyx' images mentioned above refer to ones with those regions in the center view. The 'defect' images are taken from apples with defects of various size and kind like bruise, rot, russet, scald, limb rub, hail damage, flesh damage, frost damage,...
Manual segmentation of the defected regions in 'defect' images and calyx-stem regions in 'calyx'-'stem' images were done by O. Kleynen and D. Unay, respectively. These segmentations are used as reference images to evaluate the classifier performance.

2.2. Image Segmentation

Each image contains apple skin on a darker background, so thresholding and morphological filling operation on infra-red band gives out the apple region, i.e. our region-of-interest. This is the pre-processing step of the system that is done automatically.

2.3. Feature Extraction

Equal number of healthy and non-healthy samples (pixels) were randomly selected from each image of 'defect', 'calyx', and 'stem' groups for feature extraction making a total of approximately 50000 samples for healthy, defect, calyx, and stem classes. For each of the selected samples its' color-band values as well as the averages and standard deviations of the color-band values over the whole apple area were used as feature values. Therefore, the number of features for each sample was 12.

2.4. Classifier

The classifier used in this research was a feed-forward, back-propagated perceptron network with an adaptive learning rate. It uses cross-validation technique (3/4 of the training data for training and 1/4 of it for validation) during training step and evaluates performance on test data in the testing. The following results are obtained by 'leave one out' method, i.e. sample matrices of training and testing steps do not overlap.

2.5. Post-Processing

After the training step, the classifier will classify test samples as defect or not, roughly speaking. This decision is totally sample-wise or local, so a global approach after classification is necessary to remove isolated samples as well as small defected areas. What is the limit for defected region to be small enough to be removed?
As this is an ongoing project, the details of image acquisition system are not exact. So, the size of a pixel in metric system is not reliable. That's why we selected an initial value of *50 pixels* for thresholding in post-processing step. It is applied as follows: For a test image, we take the results of the perceptron network and find the defect regions with pixel number less than the threshold and re-classify them as healthy.

2.6. Quality Grading

According to the standard of UN/ECE[1][14] concerning the international trade of apples, there are three quality classes ('Extra', 'I', and 'II') defined by skin quality and size. Maximum area of defects and minimum diameter length for classes 'I' and 'II' can be $1cm^2$ - $2.5cm^2$ and $55mm$ - $50mm$, respectively whereas the apples in 'Extra' class must be free from defects. Apples having defect region exceeding these limits are defined as 'Rejected'. At this point we introduce our defect-decision ratio as:

$$r = \frac{A_{defect}}{A_{apple}} \tag{1}$$

where A refers to area.
As a defect cannot be greater than the area of the apple skin, above ratio will always be in the range of $[0, 1]$. Using the limit values from the standard, we can reach to maximum defect-decision ratios for classes 'I' and 'II' by:

$$r = \frac{A_{defect}}{\pi \cdot (\frac{d}{2})^2} \tag{2}$$

where d refers to the diameter of the apple. So the limits of the defect-decision ratios are *0.421* and *0.1274* for classes 'I' and 'II', respectively.
As there is no clear difference in terms of limits between 'Extra' and 'I' classes in the standard and we don't have enough apples in class 'II' in our database, we will use two classes for quality of apples: 'I' (accepted with slight skin defects by ratio of *0.1274*) or 'Rejected'. The class-information of our database is in Table 1.

classes	'defect'	'healthy'	'calyx'	'stem'
'I'	146	280	145	148
'Rejected'	100	0	0	0

Table 1: Class-information of the database.

In order to easily interpret the results of classification tasks, we introduce an *Absolute Error* (ϵ) value as:

$$\epsilon = |r_{ref} - r_{seg}| \times 100 \tag{3}$$

where r_{ref} and r_{seg} refer to defect-decision ratios of reference and segmented images, respectively. With r being in the range of $[0, 1]$, ϵ will be within $[0, 100]$.

3. RESULTS

3.1. 2-Output Network Results

As an initial test we extracted defect and healthy samples from only 'defect' images and tested these images with perceptron network of 0, 5, 10, and 20 hidden neurons by 'leave one out' method. The classifier put samples into either healthy or defect classes. Post-processing was either applied or not.
Figure 2 shows images of segmentation steps, as well as the original and reference ones of a defected apple. The advantage of applying post-processing on the segmented image is obvious if we compare the last two images (i.e. segmented and segmented+post-processed).

[1] United Nations/Economic Comission For Europe

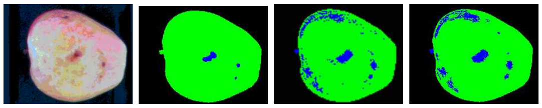

Figure 2: *Result of post-processing on an image. Left to right: original, reference, segmented, and segmented+post-processed.*

	post-processing			
	no		yes	
neuron#	ϵ	%	ϵ	%
0	19.6±13.6	43.1	17.0±13.6	50.4
5	19.9±15.9	47.2	19.0±15.8	48.8
10	19.3±16.1	50.0	18.6±16.0	49.6
20	18.0±16.0	54.9	17.4±16.0	53.3

Table 2: Results by 2-output network.

In Table 2 we see the absolute error values in the form of *average* ± *standard deviation* and average correct classification rates for these tests. As we apply post-processing the absolute errors decrease, however there are some exceptions in classification rates. If the number of hidden neurons increases classification rates also increase with an exception in '0 neuron+post-processing' case. The highest classification is observed in '20 neurons+no post-processing' case with 55 %. Comparison of resulting images with the reference ones revealed that this 2-output network tends to classify calyx and stem regions as defected, which highly affects the resulting defect-decision ratios and classifications. A 3-output (healthy, defect, and calyx/stem) network should solve this problem.

3.2. 3-Output Network Results

We created feature matrix by extracting defected, healthy, calyx, and stem samples from all the images of the database. Then for each image we trained the system with features of samples extracted from the rest of the images and tested the system with that image('leave one out' method). Post-processing was either applied or not. In Figure 3 we observe the original, reference and segmented images of four apples of database. In the virtual images black, green, red and blue colors refer to background, healthy, stem/calyx and defected regions, respectively. Stem or calyx regions are truely classified in all of them. Our system tends to classify the regions near stem or calyx as defected and some healthy skin closer to the edges are misclassified as defected.

The absolute errors and correct classification rates of image groups with and without post-processing are in Table 3. Here testing was done with all the database. The classifier used was the same perceptron network with 5

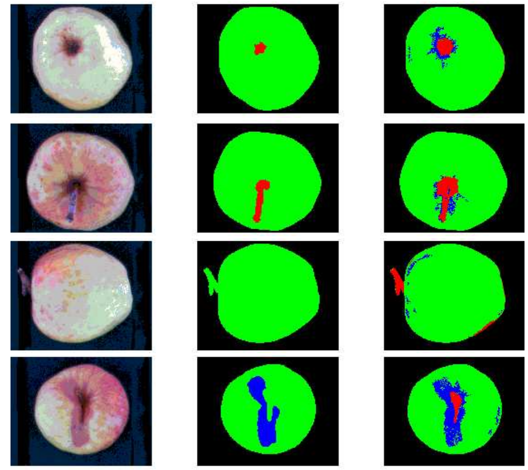

Figure 3: *Images of 3-output network with post-processing applied. Top to bottom: calyx, stem, healthy and defect images. Left to right: original rgb, reference and segmented.*

	post-processing			
	no		yes	
image group	ϵ	%	ϵ	%
'defect'	15.4±13.5	69.1	13.9±13.5	74.0
'healthy'	14.9±10.6	54.6	11.7±10.4	68.6
'calyx'	9.5±9.0	78.6	7.4±8.7	83.5
'stem'	13.2±9.5	58.8	10.6±9.7	71.6

Table 3: Results by 3-output network.

hidden neurons. If we compare the results of 'defect' group with the ones of previous test in Table 2, absolute error values and recognition rates prove that 3-output network performs better with an increase in classification rates of more than 22%. Moreover, as we apply post-processing absolute errors strictly decrease and classification rates increase for all the image groups, which shows that post-processing is necessary. However further research for the optimum threshold value of post-processing is necessary. Misclassifications in the 'stem' and 'calyx' groups were due to the regions around the stem or calyx that had darker colors. Some defect types (like 'other') were recognized as stem or calyx regions leading to misclassifications. We observe a low rate of correct classification in 'healthy' images, which is probably because our classifier is too sensitive to differences on the apple skin.

In Table 4, confusion matrix and classification rates of the above test with post-processing for the database are observed. We reached to 70.5 % and 94 % correct classification rates for classes 'I' and 'Rejected', respectively and a global rate of 73.4 %, which is slightly higher than the performance of the system Leemans[3] introduced. More-

graded in	true classes 'I' (accepted)	'Rejected'
'I' (accepted)	507	6
'Rejected'	212	94
correct %	70.5	94.0
global correct %	73.4	

Table 4: Confusion matrix of 3-output network with post-processing applied.

over the number of images tested for our system were 819, whereas it was less than 100 for his system.

4. CONCLUSIONS

We introduced some preliminary results of an apple grading system using pixel-based neural network classifier. Results of the classifier are processed and then a *defected* or *healthy* decision is made. The results of 2-output and 3-output systems are introduced and the necessity of a 3-output system for apple grading is proven. Moreover it is shown that, a post-processing step increased the classification performances. We reached to slightly better classification rates than those of Leemans[3] for a test set of 819 images, which is higher than that of his system.
There are still more to do at this point. The results shown here are obtained by a perceptron network of 5 hidden neurons, so the effect of more hidden neurons should be examined as well as other type of classifiers. We should also find the optimum threshold value for the post-processing step. Another future work is to use a priori information of the relation between stem, calyx, and defected regions on the apple skin and apply a more knowledge-based post-processing. The aim of the project is to classify apples with a speed of more than 10 fruits per second.

5. ACKNOWLEDGEMENTS

We would like to thank the providers of the image database, Prof. M.-F. Destain and O. Kleynen from Laboratoire de Mécanique Agricole of Faculté Universitaire des Sciences Agronomiques de Gembloux. We also thank O. Kleynen for originals and manual segmentations of the database.
This project(CAPA) is funded by the Ministère de la Régionne wallonne, Belgium.

6. REFERENCES

[1] V. Leemans and et al., "Defect segmentation on golden delicious apples by using color machine vision," *Comp. Elec. Agric.*, pp. 117–130, 1998.

[2] V. Leemans and et al., "Defect segmentation on jonagold apples using color machine vision and bayesian classification method," *Comp. Elec. Agric.*, pp. 43–53, 1999.

[3] V. Leemans and M.-F. Destain, "A real-time grading method of apples based on features extracted from defects," *J. Food Engineering*, pp. 83–89, 2004.

[4] K. Nakano, "Application of neural networks to the color grading of apples," *Comp. Elec. Agric.*, pp. 105–116, 1998.

[5] G. Rennick and et al., "Machine grading and blemish detection in apples," in *Proc. of the 5th Int. Symp. on Signal Processing and Applications*, 1999.

[6] W.M. Miller and et al., "Pattern recognition models for spectral reflectance evaluation of apple blemishes," *Postharvest Bio. and Tech.*, pp. 14–20, 1998.

[7] Q. Yang and J.A. Marchant, "Accurate blemish detection with active contour models," *Comp. Elec. Agric.*, pp. 77–89, 1996.

[8] Z. Wen and Y. Tao, "Building rule-based machine vision system for defect inspection on apple sorting and packing lines," *Expert Sys. App.*, pp. 307–313, 1999.

[9] F. Pla and et al., "An integral automation of industrial fruit and vegetable sorting by machine vision," in *Proc. of the 8th IEEE Int. Conf. on Emerg. Tech. Fact. Autom.*, Antibes, 2001.

[10] D. Unay and B. Gosselin, "Apple defect detection with mlp-neural networks," in *Proc. of the 13th Workshop on Circuits, Systems and Signal Processing*, Veldhoven, Nov. 2002.

[11] D. Unay and B. Gosselin, "A study on quality grading of jonagold apples," in *Proc. of the IEEE Int. Symp. on Signal Processing and Information Technology*, Darmstadt, Dec. 2003.

[12] R.O. Duda and P.E. Hart, "Pattern classification and scene analysis," 1973.

[13] O. Kleynen and et al., "Selection of the most efficient wavelength bands for 'jonagold' apple sorting," *Postharvest Bio. and Tech.*, pp. 221–232, 2002.

[14] 56th Session on Standardization of Perishable Produce and Quality Development, "United nations/economic commission for europe standard for apples and pears," 2000.