Video grading of oranges in real-time

Michael Recce(1), Alessio Plebe(2), John Taylor(3) and Giuseppe Tropiano(2)

Abstract

We describe a novel system for grading oranges into three quality bands, according to their surface characteristics. This processing operation is currently the only non-automated step in citrus packing houses. The system must handle fruit with a wide range of size (55-100mm), shape (spherical to highly eccentric), surface coloration and defect markings. Furthermore, the point of stem attachment (the calyx) must be recognised in order to distinguish it from defects. A neural network classifier on rotation invariant transformations (Zernike moments) is used to recognise radial colour variation, that is shown to be a reliable signature of the stem region.

This application requires both high throughput (5-10 oranges per second) and complex pattern recognition. Three separate algorithmic components are used to achieve this, together with state-of-the-art processing hardware and novel mechanical design. The grading is achieved by simultaneously imaging the fruit from six orthogonal directions as they are propelled through an inspection chamber. In the first stage processing colour histograms from each view of an orange are analysed using a neural network based classifier. Views that may contain defects are further analysed in the second stage using five independent masks and a neural network classifier. The computationally expensive stem detection process is then applied to a small fraction of the collected images. The succession of oranges constitute a pipeline, and, time saved in the processing of defect free oranges is used to provide additional time for other oranges. Initial results are presented from a performance analysis of this system.

Affiliation of authors

(1) Department of Computer and Information Science
New Jersay Institute of Technology
University Heights
Newark, New Jersey
email: recce@cis.njit.edu

(2) Agricultural Industrial Development (AID)
Industrial Park, Blocco Palma I
Catania, Sicily
email: aid@cres.it

(3) Department of Anatomy and Developmental Biology
University College London
Gower Street, London WC1E 6BT
U.K.
email: j.taylor@anat.ucl.ac.uk

(3) This work was partially supported by the European Community CAPRI Project, Subproject PREFER, of the High Performance Computing and Network Program

1 Introduction

Most of the processing of fresh fruit in packing houses is highly automated. Machines are used very effectively for operations such as washing, waxing, sorting by size and/or colour, and packing. However the most important step in the process, namely inspection and grading in quality, is still, with very few exceptions, performed manually throughout the world.

This step has not yet been fully automated in production systems, as it requires fast and complex image analysis. Although grading of fruit shares several common features with more classical automated inspection of manufactured goods, this problem is significantly more difficult due to the wider range in variation found in natural products. Automation of the grading process is expected to reduce the cost of this important step and to lead towards the standardisation of grades of fruit that is desired by international markets. The inspection process must grade individual oranges into a small number of qualities (typically three). Oranges are assessed according to surface characteristics including: discoloration, bruising and other blemishes. The grade is a measure of the number and size of these surface marks. When presented to an imaging system, the stem is very hard to distinguish from defects. Part of the difficulty in correctly grading oranges is the process of detecting and distinguishing the stem from other potential marks that determine the appropriate grade.

AID first developed a grading machine in 1987. Other manufacturers have also introduced machines described as orange graders [1], but none of these incorporates a stem detection mechanism, and the binning of the fruit is quite coarse. So far none of the existing machines is able to match the requirements set by the market. We know of only one prior description of an orange grading machine [1], but image processing techniques have been used to grade a wide range of fruit and vegetables (see [2] for a review).

This paper presents a new approach, largely based on the use of neural networks to achieve a more thorough analysis of the surface of oranges, including the detection of the stem. This analysis is aided by a new mechanical design which provides simultaneous images of the fruit from six orthogonal directions. More traditionally the grading decision is baded on one or more views from a single direction as the inspected object is moved passed a fixed camera [3], [4]. One of the key aims addressed by this video grading (VG) machine is the real-time performance required by the application. The detailed video-based analysis is computationally demanding, and the problem is compounded by the high throughput required for a commercial grading machine.

The key to the solution is the use of a flexible process, in which a very fast global inspection is applied by default to every view of each orange. Further, more detailed and time-consuming analyses are issued only when required as a result of the previous stages. The consequence is that the processing time becomes highly variable. In general, the computation is faster for high quality oranges, and scales roughly with the fruit size. To take advantage of this time variation, the grading decision and activation of the binning devices is delayed so that processing time can be best apportioned between the stored images of fruit currently in the pipeline. Since the largest percentage of fruit is of good quality, the time frame between adjacent oranges can be much smaller then the actual time spent to perform the most detailed analysis for a worst-case orange. The effect of this is to improve the throughput and overall grading performance (proportion of oranges correctly graded). However, correct grading is not guaranteed for many consecutive oranges if all exhibit difficult surface features.

Even with *intelligent* time allocation, it is necessary to incorporate state-of-the-art hardware to achieve the target performance at a viable cost for a commercial machine. This includes a conventional digital signal processor and a new specialised neural network parallel processor currently under development at Philips Electronics Laboratories. The design of this hardware has been described previously [5].

All of the different layers of analysis used by the complete algorithm are based on neural networks, fed by different feature extractors. This application is particularly well suited for neural network based techniques, because it is very hard to define any geometrical or spectral properties for the natural surface characteristics of fruit (stem, discoloration, bruising and other blemishes). However, the human eye is able to distinguish between a good orange, a defect, or a stem quite easily, which shows that the visual appearance, at modest resolution, contains the necessary information for classification.

2 System Overview

The *VG* automated fruit inspection machine consists of three principal parts: a system to convey the fruits into the vision chamber, the vision chamber itself, along with the image processing system, and finally the mechanism which diverts the oranges into different lanes according to their estimated quality (Figure 1). Although the most important part of the system is the image processing system, the method used to obtain the images is not at all trivial, and its development has been closely linked to the processing algorithms. The basic problem is to inspect, at high speed, the complete surface of a nearly spherical object. The wide range of sizes and shapes found in oranges preclude the use of conventional conveyors (rollers and belts). A novel design has been developed, permitting the simultaneous inspection of the entire surface of an orange during a short flight through the vision chamber.

— Figure 1 here —

2.1 Impeller

The first stage of this new system is a mechanism for accelerating the oranges to a suitable velocity, spacing them apart from each other, and propelling them along a fixed trajectory through the inspection chamber. The idea here is that each orange should follow roughly the same path independent of its geometrical properties. This should occur as long as the contribution of air resistance is minimal, the orange is not rotating when it leaves the impeller and there is good control of all of the components of its initial velocity.

Also only one orange can be in the vision chamber at any point in time (otherwise it is impossible to obtain the views in front of and behind the orange). At prior stages in the packing house the oranges flow with a high packing density (approximately 100 mm centre-to-centre separation) at roughly ten oranges per second in an individual line. With a multi-step acceleration process we increase the speed of the oranges by a factor of four and simultaneously increase the spacing by the same amount. The orange leaves the impeller with a speed of approximately 4 metres per second, travels 250mm through the air, and lands on a belt that is moving with the same speed as the horizontal component of the initial velocity. The small change in velocity experienced by the orange minimises the chances of bruising. The six views of each orange are captured simultaneously at the centre of the vision chamber.

2.2 Image capture and processing

The system hardware is VME-bus based, with four basic components:(1) the master processor (MC68030), (2) the colour frame grabber based on a Texas Instruments TMS320C40 DSP, (3) the Philips prototype board based on the L-Neuro2 parallel neural engine, and (4) an industrial digital interface board for synchronisation and actuator commands.

The amount of processing required to make a decision on the quality of an orange depends on the quantity, type and distribution of defects. As described below a perfect orange can be processed very quickly, while a complex arrangement of defects could require a considerably longer time to process. For this reason we have constructed a multi-stage processing sequence, in which decisions that require the least time are performed first, which effectively reduces the processing required at later stages. Furthermore, as described below the grading decision for an orange is delayed so that time saved in processing one orange can be applied to provide more time to process a subsequent orange.

There are three identified stages in the grading decision (identified below as separate processes). A fourth process provides the global supervision and control of the other three processes. The controlling process executes on the master processor (MC68030), supervises the other three processes, and keeps track of the location of individual oranges that are currently within the machine. The image processing processes are executed by the DSP and the L-Neuro2 (see [5] for a description of the L-Neuro2).

—-Figure 2 here—-

In order to image the entire surface of the orange, one strategy is to use six planar views normal to the axes of a Cartesian coordinate system located at the centre of the roughly spherical orange. The entire surface could be viewed, in theory, using two projections, but this results in considerable distortion from the curvature of the orange. If we assume that the orange is spherical than with six views the angle between the surface normal vector and the image plane is less than 45 degrees for over 95% of the surface of the orange. Six views are also quite practical to implement.

In the current prototype machine the six views are captured by projecting the image from a set of mirrors to an asynchronously triggered CCD camera. The camera is positioned directly above the centre of the vision chamber, and the image capture is triggered when the orange is in mid flight in the centre of the chamber. From the camera view point, four of the mirrors are in a "X" pattern and project four side views upwards. Two mirrors are used to project the underside of the orange upwards, and two mirrors are used to project the top view of the orange into the camera image. Figure 2 shows the camera view of the projected images. The image of the orange is the same size in each of the four side views, is slightly smaller (17%) in the underside view and is slightly larger (13%) in the top view. The figure also shows the region extracted from the six simultaneous views. The image chamber is uniformly illuminated with high speed fluorescent lights, but the chamber is designed so that light bulbs are not visible in any of the projected images.

In prior machines, which used a single image plane, the oranges were rolled and moved under the camera. In this case, the amount of surface seen, and the number of times that a surface element was seen, depended on the radius and eccentricity of the orange.

As soon as the image has been captured, the unique pixels are extracted from each of the six views of the orange. The symmetry of the viewing arrangement permits a calculation of the most appropriate regions from the six (assumed circular) views of the orange. The shape could be constructed by projecting the edges of a cube on to a circumscribed spherical screen, starting at a point at the center of the sphere, and then projecting this three dimensional curved cube onto a plane that is parallel to a face of the original cube. In addition, for each point selected from the image, a weighting factor is calculated which may be used to compensate for the viewing angle. Therefore, since points towards the edge of the orange are viewed more obliquely, a weighting value greater than 1.0 (ranging up to approximately 1.6) is used to give each part of the orange surface equal value in calculations of the surface intensity histogram. The initial image extraction is performed by the C40 DSP.

Afterwards, each of the views is examined, and a view is excluded from further processing if there is no chance that the particular view contains either defects or the stem of the orange. This decision is performed by a neural network algorithm that classifies colour histograms (normalised red and green) of the pixels contained in the view (described in more detail below). The DSP then passes each of the remaining views to the L-Neuro board for a more detailed analysis of local areas of the surface.

The image is scanned using a region based operator and each region is classified as defect free or not, using a second neural network (described below). The fraction of the surface area of the orange containing defects is the essential basis for the grading decision. Before treating an identified region as a defect, it is necessary to further check if it could be a stem. This is only done if the identified region is large enough to potentially be the stem. This is the most complex process in the grading operation and is carried out using a third neural network.

2.2.1 Sorting system

After the images of the orange have been obtained and the grading processing completed, the orange is deflected to the appropriate bin using purpose built pneumatic valves that release shaped jets of compressed air across the direction of motion and deflect the fruit into three separate areas.

During the visual analysis of each fruit, its dimensions, and hence mass are measured, which is used to set the timing and shape of the air jets. This activation is delayed to allow more fruits to be in a pipeline of process at the same time, and thus take advantage of the time-varying processing for each one.

3 Histogram Analysis

The first stage in the image processing is the classification of the colour distribution of each of the views of an orange. Histograms of the normalised pixel values are constructed for two of the three colour planes (red and green). Prior to this step, the colour vectors are normalised to remove the effect of slight variations in the lighting. From empirical evidence, views of oranges that do not contain a stem or defects are found to have normally distributed red and green colour components. This is to be expected, since the natural skin pigmentation is composed mostly of one colour. Although, the peak colour and its spread vary over different varieties of oranges, as well as over the season for the same variety.

Stems, discoloration, bruising and other blemishes tend to corrupt the smooth normal distribution of both the red and green pixel values. Unfortunately, this effect is often quite small, and the number of pixels with abnormal colours is unpredictable. Furthermore, in some cases a pigmentation gradient is present in good oranges, sometimes caused by non-uniform exposure to sunlight during the ripening process.

Therefore, in part the classification of a view can be computed by the fit of the frequency distribution of the pixel values in the red and green colour planes to a Gaussian function. Data which fit well to a Gaussian function are not likely to contain defects or the stem. If x is the pixel value of a colour (red or green), $h(x)$ the histogram distribution function, then μ and σ are the mean and standard deviation computed from the distribution. The best fit Gaussian function $g(x)$ is given by:

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

and the error is the summed difference between the best fit function and the histogram data. Figure 3 shows example best fit histograms for two defect free views, one view that contains a stem and one view that contains a defect. The examples in this figure show that while there is usually a good fit in the defect free images, there is a degree of variation. Also the view containing a stem and the view containing a defect can not be distinguished using the histogram analysis.

—Figure 3 here—-

The surface defects on the oranges produce characteristic ranges of error in specific segments of the histogram. For this analysis, the histogram data is divided into a set of segments I_i, where:

$$I_i = |\sum_x h(x) - g(x)|_{\mathbf{x} \in \mathcal{X}_i}.$$

Empirically we found that a simple scoring rule based on the fit of a Gaussian function to the original histogram does not perform as well as a neural network based classifier that is used to learn the characteristic differences in a measured distribution from a normal distribution. The neural network algorithm used here is a modified form of back-propagation training of a multilayer perceptron [6][7] with ten inputs and two outputs. The input layer combines information from the red and green pixel colour component histograms. There are two output classes: (1) no defects, or (2) either defective or containing a stem. Views of oranges that fall into the second class are passed to the second stage of the process, in which masks are used to analyse the surface in more detail. The number of input neurons, the partitioning of the histograms, and the number of hidden units have been empirically evaluated and compared.

As with other classification procedures, this procedure incorrectly classifies a fraction of the top quality oranges (Quality I) into a lower quality band (Quality II and Quality III), and places a fraction of the lower quality oranges into the top quality band. From the commercial point of view the full system should make fewer errors in downgrading the quality than in upgrading the quality. However, in this first stage of the processing it is safer bias the process towards downgrading, and to run the additional processing stages when there is the possibility that a view of the orange contains a defect.

As suggested by Hush and Horne [8], better generalisation is achieved if the number of training samples is at least ten times larger then the number of weights in a multilayer back propagation network. This neural network, and the neural networks at later stages of processing, is trained from a database that contains over 2000 examples (including a 20% test set).

4 Local Defect Search

One approach to the identification of defects is to segment the defect regions from the parts of the surface that are defect free. A detailed analysis of the characteristics of defects leads directly to the observation that there is no regular clustering of defects in colour space. Also in several cases, local colour variation between pixels from the same orange is not sufficient to classify defects. The multiplicity of defect types has hampered prior attempts to segment the defects from the clear surface regions, using local structural or textural properties. Furthermore, it is not clear that the computational effort required to identify all possible classes of defects, if it were possible, is justified, since all defect types contribute roughly equally to the final grading decision.

Instead we construct a simple and effective defect detector using a set of masks applied to regions of the image. The extracted local features are passed to a neural network based classifier that has been trained using a large database of samples. In almost all cases, a defect is characterised by a discontinuity in the skin pigmentation. But the colour gradient can nevertheless be quite small, and similar to other non-defect sources of variation such as uneven illumination and border effects.

4.1 Defect feature operators

Ordinary edge detectors do not respond strongly to the appearance of surface defects on oranges. Instead, five larger and lower spatial frequency operators were used. In this process we divide the orange in regions (NxN), of the typical size of defects, and apply all five operators to both the red and the green colour planes. Each operator (M_k) is an NxN matrix, with integer elements that are either one, zero or minus one. The elements (m_{ij}^k) of the first four of these matrices are defined below.

$$m_{ij}^1 = \begin{cases} 1, & i < n/2 \\ -1, & otherwise \end{cases} \quad m_{ij}^2 = \begin{cases} 1, & j < n/2 \\ -1, & otherwise \end{cases}$$

$$m_{ij}^3 = \begin{cases} 1, & i > j \\ 0, & i = j \\ -1, & otherwise \end{cases} \quad m_{ij}^4 = \begin{cases} 1, & i < (N-j) \\ 0, & i = (N-j) \\ -1, & otherwise \end{cases}$$

In order to specify the elements of the fifth matrix (M_5) it is useful to define M_k', with elements $m_{ij}^{k\prime}$, to be constructed in the same way as (M_k) except that $N' = N/2$. M_5 is composed of four quadrants which are defined to be:

$$m_{ij}^5 = \begin{pmatrix} -m_{ij}^{4\prime} & -m_{ij}^{3\prime} \\ m_{ij}^{3\prime} & m_{ij}^{4\prime} \end{pmatrix}$$

The image is divided into regions, each of size NxN, and $x_{ij}^{r,g}$ are the red and the green components of the pixels in a region. The neural network based classifier is applied to each of the image regions. The input layer of the neural network contains ten neurons, computed by separately applying each of the five masks to the red and to the green components.

$$I_k^{r,g} = \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ij}^{r,g} m_{ij}^k$$

The output layer of the neural network contains two neurons, and they are trained to classify defects, and used to estimate the severity of the defect.

—Figures 4 and 5 here—-

The data base used to train this neural network based classifier is the same as the one described above. However in this case the parts of the image that correspond to defects or the stem have been extracted. Figure 4 contains part of the database of defects. Note that there are several examples of the same defect, which differ only in the selection of the start point for the sampled region. In order to reduce the computational overhead for a single view, we did not convolve the entire view with the defect detection masks. The masks were applied only to a coarser tesselation of the image. With this constraint, the performance of the classification is significantly improved if the neural network classifier is trained on images that have been translated in both directions.

Figure 5 contains samples of the stem data base. In this part of the analysis the stems are also classified as defects. This data base is then used in the next stage of processing to differentiate between the stems and defects. This stem detection process is only used if the stem has not already been identified in the prior processing of the six views of an orange.

5 Stem Detection

Stem detection is the most difficult step in the process of grading oranges, since the stem region looks very similar to a defect. In general, global features such as the average colour or the distribution histogram of red and green pixel colour components, do not show significant differences between stems and defects. The strategy chosen here is to use a neural network based classifier, applied to identify spatial features extracted from the image region that contains the suspected stem.

The stem has a much more regular structure than the defects and has a high degree of radial symmetry, which can be used to aid the identification of the stem. Often, but not always, there are characteristic radial groves that radiate from the stem attachment area. Also in some cases there are concentric rings around the stem attachment area. Unfortunately circular defects, of various types, also occur relatively often.

We have found that the family of Zernike moments are very useful in detecting stems. In this process a set of Zernike moment masks are convolved with the neighbourhood of defects, identified in the prior step, in order to search for evidence of a stem.

—Figure 6 here—

5.1 Zernike moments

When used as a convolution mask, the family of Zernike moments is particularly sensitive to circular symmetries. Importantly, Zernike moments are invariant under rotation, which makes this strategy computationally practical. Each of the views of an orange processed by this algorihm has already been scanned for defects using masks much smaller than the Zernike moment mask, so this stage is quite specific to suspected stem regions.

In these tests the radial size of the Zernike moments was not varied, since the size of stems of each particular variety of oranges is relatively fixed. However, for a commercially viable machine, in which a large range of fruit sizes are processed, some scaling of the mask size may be necessary.

In order to generate a two dimensional square mask based on a Zernike polynomial, the mask is taken to have a unit radius in polar coordinates. Each pixel is assigned a complex value:

$$Z_{nm}(\theta, r) = [\cos(m\theta) + j * \sin(m\theta)] * R_{nm}(r)$$

where: n = the major order of the Zernike polynomial; m = the minor order of the Zernike polynomial; and R_{nm} = the radial part of the Zernike polynomial. The Zernike polynomial R_{nm} is a sum, from $s = 0$ to $(n - m)/2$ of the following terms:

$$R_{nm} = \sum_{s=0}^{s=\frac{(n-m)}{2}} \frac{(-1)^s \ (n-s)!}{s! \ (\frac{n+m}{2} - 2)! \ (\frac{n-m}{2} - 2)!} \ r^{(n-2s)}$$

The radial variation is given by the R_{nm} polynomial, which has degree N, with only terms of the same parity as N. It generate an oscillating function with maximum wavenumber $\frac{1}{4}N$. The minor order also affects the radial component, weakeNing the amplitude of the oscillations in the proximity of the circle center, while enlarging its wavelengths. The border oscillations have larger amplitude and shorter wavelength as M approaches N .

Figure 6 shows three examples of Zernike masks, represented in the picture with the intensity given by one component ($Img(Z_{9,5})$; $Real(Z_{16,8})$; and $Mod(Z_{5,3})$). More details on Zernike polynomials and their use as masks for image processing can be found in Khotanzad and Hong [9].

In these masks, higher frequencies, both in radial and angular directions, vanish inside the mask. This occurs particularly near the centre of the mask, where quantisation effects are largest. Also the size of the window must be large enough to fully enclose the stem region. Reliable detection is only possible when the mask is well centred with respect to the stem region.

With the Zernike moments, the best input features for a neural network cannot be found empirically. For each Zernike polynomial major order n the possible minor orders are:

$$m = \begin{cases} n/2 + 1, neven \\ \\ (n+1)/2, nodd \end{cases}$$

Limiting the polynomial orders as mentioned above, to some N_{MAX}, the number of candidate features is:

$$F = 6\ \times \sum_{N=2}^{N=N_{MAX}} N \times \#M$$

The factor 6 is the number of additional combinations that each Z_{nm} requires when applied to the two colours: red and green, and in all three possible components ($Real(Z_{mn})$, $Img(Z_{mn})$ and $Mod(Z_{mn})$).

From this huge feature space, a smaller set, which results in the best classification, must be found. It is impossible to find the best set by evaluating all combinations. Let F^* be the largest feature dimension to be used by the network. Exhaustive optimisation would require:

$$\sum_{f=1}^{f=F^*} \frac{F!}{(F-f)!}$$

comparisons, each over a database of samples. For example, to extract the best feature space when $N_{MAX} = 30$ and $F^* = 5$ over a database of 5000 samples, would require billions of years on a SUN Sparc20 workstation.

It was decided to run an exhaustive evaluation to determine the best set over a smaller number of candidate features. The selection of the four input neurons, used in this classifier, from 50 candidates required 200,300 runs against the database, taking approximately 100 hours.

The screening of the subset of candidates is done by ranking all the features with a single-variate analysis. For each feature i the rank is computed as:

$$\frac{R_i = |m_{i,s} - m_{i,d}|}{s_{i,s} + s_{i,d}}$$

where:

$$m = E[x]; s = E[(x - m)^2]; \qquad {\scriptstyle i=feature\ index,} \begin{cases} {\scriptstyle s,} & \text{stem class} \\ {\scriptstyle d,} & \text{defect class} \end{cases}$$

and x is a single feature value.

During the exhaustive optimisation, ranking is done in the multivariate space. Classification during the analysis and test is made according to the minimum of the distances:

$$dist(\mathbf{x_i}, \mathbf{m_i}); \quad {\scriptstyle i=s,d}$$

The table shows the rank of the example masks presented in Figure 6.

Zernike mask	colour plane	score
$\mathbf{Img}(\mathbf{Z_{9,5}})$	red	14.8
$\mathbf{Img}(\mathbf{Z_{9,5}})$	green	25.1
$\mathbf{Real}(\mathbf{Z_{16,8}})$	red	53.3
$\mathbf{Real}(\mathbf{Z_{16,8}})$	green	16.8
$\mathbf{Mod}(\mathbf{Z_{5,3}})$	red	38.4
$\mathbf{Mod}(\mathbf{Z_{5,3}})$	green	44.8

6 Results

The research presented in this paper is aimed directly for use in a commercial machine with stringent real-time requirements. Any envisaged solution should be cost-effective, robust, simple, and suited for implementation in parallel hardware. Nevertheless, the problem of inspecting fruits is intrinsically complex, and previous attempts to simplify it have been at best partially successful.

The flexible architecture of the algorithm presented here provides an advantage, by splitting the whole task into three simpler problems. The use of neural networks in solving all the steps of the grading has overcome some of the difficulties in defining a reliable description of the properties of the objects to be recognised and classified, which are intrinsically difficult to specify, and the results are so far within the expected performances of the *VG* machine.

—– Figure 7 here —

Figure 7 shows example results from four views of oranges. One of these views contains no defects, two of the views contain defects which have been detected, and the fourth view contains a detected stem. The following table summarises the errors that resulted in a test using the highest performance version of the neural network algorithms tested thus far.

Network	Training Set		Test Set	
	→ down	→ over	→ down	→ over
Histogram	0.00%	0.00%	13.73%	2.06%
Defect	0.00%	0.62%	0.80%	2.02%
Stem	0.11%	0.87%	0.44%	1.31%

The database was constructed using 200 oranges of the tarocco variety, and included 2296 images of stems and 2567 images of defects. Eighty percent of the database was used for training and twenty percent was used to evaluate the performance of the algorithms. The error labelled "→ down" indicates an incorrect down grading of one of the views of an orange. For the histogram evaluation neural network, this means that further analysis is performed on an orange that is defect free. When the neural network used to detect defects erroneously down grades an orange than the region is searched for a stem. If the stem detection down grades the view then the stem is erroneously taken to be a defect. The errors listed in the table in the "→ over" columns are false positives, in which defects and stems are passed undetected.

The imbalance in the two error classes is intentional, as described above. Moreover the differences in performance between the three neural network algorithms reflects the expected relative significance of errors in the final decision. It seems clear that the histogram analysis, which contains no spatial information, is unlikely to be refined significantly. Nevertheless, this is the processing step with the largest difference between the training and test errors, reflecting a rather poor level of generalisation. It may be that in this case the training set is not large enough. For a given number of oranges in the database, the histogram training set provides a smaller number of samples compared with the other two methods.

Evidence for the advantages provided by use of neural networks is shown here for the stem detection problem, which is the most important part of the whole algorithm. As a two class problem, the classical Fisher's linear discriminant seems a promising candidate technique[10]. When this technique is applied to the sample database, using the same reduced feature number used in the neural network results presented above, the results are quite poor. In this case 27.5% of the stems are not detected and 45.8% of the stems are mis-classified as defects.

This linear method may fail due to the implicit assumption that the distribution function of each feature is normal, and this is not true in the case of the Zernike convolutions over the database. In comparison, the neural network algorithm does not explicitly assume the characteristics of the probability distribution of the samples.

7 Summary

The present algorithm for fruit inspection, based on neural network classifiers, has come close to meeting the requirements of a future commercial video grading machine. Further work is envisaged to refine all stages of the classifier, and to balance the computational requirements of each of the stages. Also, thus far, little attempt has been made to use network architectures other than back-propagation learning in a multi-layer perceptron. We plan to compare the performance or alternative non-linear classifiers. The present choice of approach has also been guided by the efficiency of running multi-layer perceptron algorithms on the currently available hardware. The machine has been tested with a a few hundred oranges and a range of varieties. However, the machine may require a degree of profiling in order to cope with the large set of varieties of oranges that are commercially available. For example, the machine may need to be adjusted for particularly small or particularly large oranges. Also the machine may require the development of specialised sets of synaptic weights and other parameters in order to optimally grade particular varieties of oranges.

Acknowledgements

The authors would like to thank Nick Ouroussoff and Dave Edwards for invaluable contributions to the design and construction of the video grading machine.

References

[1] H. Maeda. Grade and colour sorter through image processing. In *Proc. of Intern. Symp. on Agric. Machinery and Intern. Coop. in High Tech. Era*, pages 350–356. Univeristy of Tokyo, 1987.

[2] R.D. Tillett. Image analysis for agricultural processes: a review of potential opportunities. *J. aric. Engng Res.*, pages 247–258, 1991.

[3] A. Davenel, C.H. Guizard, T. Labarre, and F. Sevila. Automatic detection of surface defects on fruit using a vision system. *J. aric. Engng Res.*, pages 1–9, 1988.

[4] S.A. Shearer and F.A. Payne. Color and defect sorting of bell peppers using a machine vision system. *Trans. ASAE*, pages 1–9, 1988.

[5] N. Maudit, M. Duranton, J. Gobert, and J.A. Sirat. Lneuro1.0: a piece of hardware lego for building neural network systems. *IEEE Trans. Neural Networks*, pages 414–422, 1992.

[6] D.E. Rumelhart, G.E. Hinton, and R.J. Williams. Learning internal representations by back-propagating errors. *Nature*, 323:533–536, 1986.

[7] P.J. Werbos. *Beyond regression: new tools for prediction and analysis in the behavioral sciences.* PhD thesis, Harvard University, 1974.

[8] D.R. Hush and B.G. Horne. Progress in supervised neural networks -what's new since Lippmann? *IEEE Signal Processing*, pages 8–39, 1993.

[9] A. Khotanzad and Y.H. Hong. Invariant image recognition by zernike moments. *IEEE Trans. pattern analysis and machine intelligence*, pages 489–497, 1990.

[10] R.O. Duda and P.E. Hart. *Pattern classification and scene analysis.* Wiley-Interscience, 1973.

[11] G. Grasso and M. Recce. Scene analysis for an orange picking robot. In *Proc. of 6th Inter. Congress for Computer Technology in Agriculture*, pages 275–280. VIAS, 1996.

Figure Captions

Figure 1 The prototype videograding machine. Part A shows oranges being carried by the impeller into the vision chamber. Note that there are two stages of acceleration, and in each of these the spacing between the oranges increases. Part B is a view of the oranges as the exit from the vision chamber and are sorted into three grades.

Figure 2 Six simultaneous views of an orange collected in the inspection chamber. The views are constructed using a set of mirrors. Part A shows an example image captured by the camera, and part B shows the regions extracted from the image. Colour is used to segment the oranges from the background [11]. In each view only the unique part is extracted as shown by the rounded shape that is drawn inside each of the circles.

Figure 3 The red and green colour distributions of four example views of oranges (solid lines), and the best fit Gaussian function (dotted line) to each of these distributions. Part A and B are the red and green pixel value distributions of a view that does not contain defects or the stem. Note that the distributions fit well to a Gaussian function. Part C and D contain the red and green colour distributions for a second example view of an orange that does not contain defects or a stem. Note that the fit to a Gaussian function is less good, and that the means and standard deviations of the two distributions are different for this example orange. Part E and F contain the colour distributions and best fit Gaussian function for an image that contains a stem and part G and H are the colour distributions for a view that contains a defect.

Figure 4 This figure contains 100 examples from the database that is used to train the mask based stem and defect detector. Note that individual defects are entered several times into the database with different displacements from the edge of the mask. The reason for this is explained in the text.

Figure 5 This figure contains 100 examples from the database of stems. This data base is used both to train the mask based defect and stem detection neural network, and is used to evaluate the performance of masks constructed from Zernike moments.

Figure 6 This figure contains three examples of Zernike moment masks. In each case the intensity corresponds to one component of the mask. Part A contains the modulus of $Z_{5,3}$, part B contains the imaginary part of $Z_{9,5}$ and part C contains the real part of $Z_{16,8}$.

Figure 7 The combined result of applying the full set of processing stages on four example views. The first view (part A) does not contain defects or a stem. The second view (part B) contains a detected stem, and the remaining views (parts C and D) contain views with detected defects. The histogram for these three of the views (A,B,D) were presented in Figure 2 (A,B,E,F,G,H).