Defect sorting of dry dates by image analysis

D. WULFSOHN1, Y. SARIG2 and R.V. ALGAZI3

1Agricultural and Bioresource Engineering Department, University of Saskatchewan, Saskatoon, SK, Canada S7N 0W0; 2Institute of Agricultural Engineering, Volcani Institute, ARO, Bet Dagan, Israel 50 250; and 3Department of Electrical and Computer Engineering, University of California, Davis, CA 95616, U.S.A. Received 9 June 1992; accepted 1 June 1993.

Wulfsohn, D., Sarig, Y. and Algazi, R.V. 1993. Defect sorting of dry dates by image analysis. Can. Agric. Engr. 35:133-139. Spectral characteristics were obtained for good quality and defective fruit of two dry date varieties, 'Majhul' and 'Zahidi'. An image processing algorithm was developed to sort good and defective fruit using simple thresholding operations. To eliminate the effects of varying system characteristics, defective date spectral characteristics were normalised based on an average 'good date' reference spectrum. Thresholding parameters were determined by integrating the normalised reflectance curves over the wavelength range in which defect/good date contrast was maximised. The reflectance data were also used to select optical filters to enhance the features of interest in the images. Preliminary results indicate that the method proposed offers a promising means of rapidly detecting defective fruit. Further research is needed to estimate sampling errors associated with the technique and to investigate real-time sorting speed.

Keywords: image processing, spectral reflectance, sorting, dates

Des caractéristiques spectrales ont été obtenues pour deux variétés, 'Majhul' et 'Zahidi', des dattes sèches, d’un part de bonnes qualités et d’autre part de qualités déficientes. Un algorithme procédures d’images a été développé, fonctionnant avec un seul unique gris, pour la classification de bons et de dattes déficientes. Pour éliminer les effets de variation des systèmes caractéristiques, les caractéristiques spectrales pour dates défectueuses sont normalisées, basées sur un spectre moyen de référence pour bonnes dattes. Les paramètres de seuil sont déterminés en intégration, sur le long des longueurs d’ondes dont le contraste entre bonne et défectueuses dattes est maximal, des courbes de réflectance normalisé. Les données de pouvoir réflecteur sont utilisées pour choisir les filtres optiques afin de rehausser les traits caractéristiques des images. Des résultats préliminaires indiquent que la méthode proposée découverte rapidement les dattes défectueuses. Il manque des recherches plus poussées pour l’estimation des erreurs d’essai et d’incertitude associées avec cette technique et pour l’étude de la vitesse de classification en temps réel.

INTRODUCTION

The date palm is native to the subtropical deserts of the Middle East and Northern Africa. Most of the present world annual production of 2 to 2.5 million tonnes from over 70 million palms is from these regions (FAO 1982). Nevertheless, the date industry in the U.S.A. is becoming of increasing importance with the larger demand for a more diversified fruit supply. Almost all U.S.A. produced dates are grown in California, amounting to 18,000 t/a with a total value of $14 million (Moyer 1988).

There are over 100 varieties of dates of which fewer than 20 are commercially important. The date fruit varies in texture, sugar type, colour, and moisture content depending on cultivar and curing treatment. Dates are classified as soft, semi-dry, and dry types. After harvesting, which is still done manually, the dates are brought to the packing-house where they are sorted based on quality (Rygg 1975). Fumigated dates are washed with water to remove foreign material and contaminants (Hussain 1974). The washed dates are graded a second time and then washed again with 0.25% ethylene oxide solution. Alternatively, in some packing houses a solution of 50% ethylene oxide is added to the packed dates.

The purpose of sorting is the removal of dates unsuitable for marketing and the separation of the remainder into several grades based upon quality criteria related to size, shape, surface colour, and visible defects, as well as marketing attributes such as moisture and sugar content which may be linked to outside appearance and texture and are also dependent on the variety of date.

The manual sorting procedure is very labour intensive, tedious work leading to inconsistent results as well as worker fatigue and line hypnosis. A comprehensive sorting machine capable of addressing both external and internal quality attributes of dates is not yet technically practical; however, the development of a mechanical sorting machine based on processing images of the fruit could significantly improve the sorting of dates and enhance the competitiveness of the industry.

PREVIOUS WORK

Size sorting of dates is a relatively simple operation and is already being done mechanically, using either an oscillating screen or a non-vibratory diverging rod system. Huxsoll and Reznik (1969) have shown that there is a correlation between shape, size, texture, and even moisture content; however, a more direct measurement is required for more reliable and consistent results. Huxsoll and Reznik (1969) found good correlation between light reflectance and visual quality attributes, but the parameters required for optimal operation were numerous and hard to control. They also investigated light transmittance which yielded promising results, primarily when using wavelengths in the near-infrared (NIR) region.

The first known, but undocumented, work on the sorting of dates using image processing techniques was performed by the 'Vartec' Co. in California, contracted by 'Tenneco' (Personal communication, P. Godzins, Golden State Machineries, Los Angeles, CA). It was concluded that it is conceptually feasible to develop a machine capable of sorting dates by processing the image of the surface. No continuation of this work is known to have taken place.
Another feasibility study was conducted in Israel to explore the possibility of implementing image processing techniques to sort dry varieties of dates (Hezroni and Fikus 1988). The investigators used a commercially available system supported by FORTH and VLC languages, which enabled fast execution of operations. Various algorithms were employed to find correlations with parameters associated with surface texture. It was concluded that it is possible to derive a mathematical model incorporating the relevant parameters for sorting into different grades.

QUALITY STANDARDS

A California standard states that quality factors of dates are freedom from dirt, decay, black scald, fermentation, and surface sugar crystallization (Ryall and Pentzer 1974). Israeli Standard 125! (SII 1984) also identifies insect infestation, exposed pit or pieces of pit, scars or stains on the skin, unripe fruit, and fruit that are too small (less than 6.5 g) as unsuitable. Dates may be soft and syrupy, or firm and dry, depending on the type and end use (Ryall and Pentzer 1974). Moisture content is also an important parameter, since the higher the moisture content, the more subject the dates are to quality loss as a result of physiological causes and attack by microorganisms. Colour is another important quality attribute. There is a broad range of colours which is considered acceptable for the different varieties of dates, although extremes of lightness or darkness or off-colours such as greenish or reddish shades are undesirable. For any given variety, the colour range is much smaller, and a lack of uniformity in colour among dates in a single package or between different packages is not desirable. Changes in colour may be the result of pigments formed during maturation or from partial dryness related to cultural practices.

OBJECTIVE OF STUDY

The objective of this investigation was to identify optical features of both marketable quality and blemished or defective dates and to use them in the development of an image processing algorithm for quality sorting.

INSTRUMENTATION AND EXPERIMENTAL PROCEDURE

Spectral reflectance measurements

The fruit used in this study was obtained from date growers in the Arava region of Israel. Diffuse spectral reflectance curves were obtained for two dry varieties of date: 'Majhul' and 'Zahidi', using a spectrophotometer (Model 330, PerkinElmer, Norwalk, CT) equipped with a 60 mm diameter integrating sphere and an 18 mm diameter viewing aperture. Illumination was provided by a Wt lamp. Reflectance below 878 nm was sensed with a photomultiplier tube, and above 878 nm with a lead sulphide detector. The instrument was calibrated using an aluminium oxide plate as the 100% reflectance standard. Measurements were made for eight good and nine defective Majhul dates, and for two good and six defective Zahidi dates over the range 400 to 1100 nm with a scan speed of 350 nm/min.

Any technique developed must be independent of illumination conditions, varietal and seasonal variations in date features, and other variable system characteristics. For this reason, some form of normalization is recommended. The approach proposed in this study is to normalize the fruit reflectance data with respect to acceptable quality or 'good' date characteristics. This will lead to a sorting procedure based on the relative characteristics of defects to average 'good' characteristics. For each variety of date, the spectral characteristics of the good quality fruit were averaged to give a reference spectrum for the given illumination conditions. The averaged spectral characteristics of each type of defect within each variety were normalized with respect to this reference spectrum:

$$R_N(\lambda) = \frac{R(\lambda)}{R_{good}(\lambda)}$$

(1)

where:

- $R(\lambda)$ = reflectance at wavelength $\lambda$, and
- $R_{good}(\lambda)$ = average reflectance of good date flesh.

The resulting curves were used to identify whether any filtering is required and in which range of RGB band differences between good and defective dates are maximized.

Imaging facilities

The image acquisition and manipulation were carried out using the image processing facility in the Department of Electrical and Computer Engineering at the University of California, Davis (Wulfsohn et al. 1989). Images were taken using a JVC KY-210 RGB video camera equipped with a Nikon 55 mm f/3.5 lens. The camera was connected to a Gould DeAnza IP8500 image processing system to digitize the images, to perform various manipulations, and to display and plot the results. Procedures and algorithms were executed interactively using software available on DAISY (Davis Interactive SYstem).

To reduce storage and processing requirements, images 280 x 350 for the 'Majhul' and 256 x 256 pixels in size for the smaller 'Zahidi' variety (instead of using higher resolution), were processed. Each of the RGB components of the resulting images had a resolution of 256 ($2^8$) grey-level intensities. For good quality images it is imperative to have good illumination and avoid shadows around the date. Specular reflectance was found to be a major problem for the dates. To minimize glare, an attempt was made to provide diffuse illumination. The sample fruit was positioned on white paper to provide a uniform, non-reflecting background having good contrast with the fruit. Illumination was provided by four 75 W, (120 V AC, soft white light bulbs shining through translucent screens) to further diffuse the incoming light. Various backgrounds were tried, including black foam to provide a truly non-glab low intensity surface. However, on this background it was not always possible to distinguish good flesh of the 'Majhul' date body or decomposing flesh on the 'Zahidi' variety, which has near zero intensity. Therefore, ali dates were digitized against the white background; nevertheless, the background material was not entirely non-reflecting, resulting in some specular reflectance about the edges of the fruit, which introduced some error into the results.

The reflectance data indicated that additional filtering would enhance the features of interest (This is discussed in
more detail in the results section). For both varieties, an infrared cutoff filter was used when taking images. Additional images were taken of the ‘Mahnl’ dates with the IR cutoff filter combined with optical filter 2.

A description of the fruit used in this stage of the study is given in Table I. The fruit used for image processing were not the same ones used in obtaining the spectral characteristics. All the dates were washed before processing.

**ALGORITHM DEVELOPMENT**

**Proposed steps**

The proposed steps for performing the image processing were:

- Digitize image
- Separate date from background (Image segmentation)
- Feature enhancement
- Classification of fruit:
  - (a) Statistical evaluation to obtain sorting criteria
  - (b) Criteria ≥ Defect Threshold

The proposed algorithm is illustrated in Fig. 1 and described in the following sections.

**Image segmentation**

The separation of the fruit body from its background is a key step in the image analysis. If the intensity of the background is distinctly different from the range of intensities associated with the date image, the image of the fruit may be easily segmented from the background using a single global threshold, \( t_0 \). Any pixel with grey level intensity less than \( t_0 \) is considered a date body pixel and is assigned a value of 1, otherwise it is a background pixel and is assigned a value of 0, that is:

\[
u(m, n) = \begin{cases} 
1 & \text{if } u_0(m, n) < t_0 \\
0 & \text{otherwise}
\end{cases}
\]  

where \( u_0(m, n) \) is the grey level of a pixel \((m, n)\) of the original digitized image of the fruit against the light background. In this way a binary image of the date to be used as a mask, \( u(m, n) \), is obtained. In future pixel manipulations, only pixels under the mask having values of 1 are processed. A logical AND operation is then performed between the binary image, \( u \), and the original digitized image, \( u_0 \). The resulting image, \( v(m, n) \) in Fig. 1, is of the fruit set against a black (0’s) background.

**Feature extraction**

Defective pixels were then determined using simple amplitude thresholding as follows: The average ratio of defective/good reflectance obtained from the normalized spectrophotometric curves in the wavelength range of interest, \([\lambda_1, \lambda_2]\), yields a corresponding scale factor:

\[
S = \int_{\lambda_1}^{\lambda_2} R_N(\lambda) \, d\lambda
\]  

A threshold, \( t \), was then computed as:

\[
t = S \times \left( \text{Average grey level intensity of ’good’ fruit} \right)
\]

\[
S = \frac{\sum_{m,n \in \text{body}} v(m,n)}{\sum_{m,n \in \text{body}} 1}
\]  

and the defective pixels obtained by simple thresholding of the image \( v \):

\[
w(m,n) = \begin{cases} 
1 & \text{if } v(m,n) > t; \quad \{m,n: (m,n) \in \text{fruit body} \} \\
0 & \text{otherwise}
\end{cases}
\]  

where the binary image \( u \) has been used as a mask to identify the body pixels. The resulting image \( w(m,n) \) is a binary image consisting of all 0’s except for the region classified as defective which are 1’s.

**Sorting procedure**

The proportion of ‘defective’ pixels was then calculated from \( w(m,n) \) as:

---

1 Kodak 301A, 0.5 density aimed at 678 nm, (Eastman Kodak Co. 1985)
2 Deep orange, cuts light below 550 nm, 0.25% transmittance at 550 nm, 60% at 570 nm.
\[
\% \text{Def} = 100 \frac{\sum \sum 1}{\text{number of pixels identified as 'defect'}}
\]
\[
= 100 \frac{\sum \sum 1}{\text{total number of pixels in date body}}
\]

This parameter may then be used as a sorting criterion.

\[
\text{If } \% \text{ Def} > 10, \text{ the fruit is classified as 'defective' else, otherwise the fruit is classified as 'good'}
\]

The parameter \(t_3\) indicates the allowable proportion of defects (including those pixels incorrectly assigned as defective) and its value would be decided by the date industry.

The proposed algorithm has potential as a very rapid means of sorting fruit. The average good fruit intensity need only be obtained once by obtaining the averaging body intensity of, say, 100 good fruit at the beginning of the day or season. This in effect calibrates the sorting machine since the normalized scale factor, \(S_n\), remains constant for a given variety. Subsequent sorting then simply involves one logical operation and two or three thresholding operations over the image area.

**RESULTS AND DISCUSSION**

**Spectral reflectance**

The averaged diffuse reflectance for the 'Majhul' and 'Zahidi' dates and the reflectance normalized with respect to the good date spectrum in the visible range are shown in Figs. 2 to 5. These results indicate that the use of an infrared cutoff filter for wavelengths between about 700-1100 nm (most cameras are not sensitive to light above this wavelength) to enhance defects on the original images even before processing may be possible, since much of the reflected energy is in this range. The Kodak 301A infrared cutoff filter was selected for this purpose. Filters are also expected to be useful to attenuate wavelengths below 600 nm (orange) in the visible range for 'Majhul' variety, and below about 460 nm and above 600 nm for the 'Zahidi' variety. The deep orange optical filter 22 was used to accomplish the former; however, no additional filters were used for the 'Zahidi' variety in this study.

The spectrophotometric results also indicated the need to digitize only the red band for the 'Majhul' variety, and the green band to detect most of the defects studied for the 'Zahidi' variety, a much smaller memory requirement than digitizing all three bands.

**Image processing**

For the 'Majhul' images, where only the IR cutoff filter was used, the red band image showed the best detail over the surface of the date, but poor contrast with the background. The blue band image gave the best fruit/background contrast but no detail over the fruit body. When the deep orange filter 22 was used along with the IR cutoff filter, the red band image showed very good detail over the date body as well as good contrast with the background (Fig. 6(a)) and was suitable for both image segmentation and feature extraction.

In the case of the 'Zahidi' variety, the green band image showed the best contrast between good and defective pixels as well as fairly good contrast with the background and was used to both produce the mask and to extract features.

The values for the threshold \(t_o\) used in these experiments are given in Table II. These were found by trial and error using interactive thresholding capabilities of the system software. Obviously, \(t_o\) could be computed in a similar manner to \(t\) (Eq. 5) based on the ratio of background reflectance to the lower limit of grey-level intensities expected for the dates. Figure 6(b) shows the mask image, \(u(m,n)\), obtained for the 'Majhul' date with the dry end.

Since for the 'Zahidi' variety, certain defects such as decay have lower intensity than good flesh, both a lower- and an
Fig. 4. Spectral characteristics of 'Zahidi' dates. Each curve is for a single fruit.

Fig. 5. Reflectance characteristics of 'Zahidi' dates normalized with respect to the good date spectrum. Each curve is for a single fruit.

upper-bound threshold ratio were obtained, $S_1$ and $S_2$, respectively. The $S_1$ value was obtained using the 'decay' spectrum, and $S_2$ was chosen as a value in-between the average values corresponding to the 'good-blister' and the 'blister' spectrums to compromise for the fact that some defects have intensities very similar to, or within the range of intensities associated with good dates. Values of thresholding scales, $S_1$, and thresholds, $t_i$, obtained are also presented in Table II.

The results of the image processing algorithm are tabulated in Table III and processed images obtained for the 'Majhul' date with dry end are shown in Fig. 6(d) and (e) for both filter configurations. The extra filtering for the 'Majhul' dates noticeably improved the results obtained; the algorithm detected a higher proportion of defective pixels, and also incorrectly assigned fewer good pixels. These results suggest we could probably improve 'Zahidi' results with appropriate filtering. The main source of error in the 'Zahidi' variety was due to the presence of some specular reflectance at the edges, which thresholding ($t_i$) categorized as defects (Fig. 7). It is expected that this problem can be minimized with better background and illumination configuration. This problem was minimal for the 'Majhul' variety with its rougher surface, and improved with filtering (Table III). As expected from its spectral characteristics, red dryness of the skin could not be detected using this technique, since the intensity values were within the range of good reflectance (Fig. 7(d)).

FURTHER CONSIDERATIONS

In the method described in this report, a colour camera was used. For the purposes of industrial implementation of a sorting machine, cost is always an important factor. The results of these experiments suggest that a B&W camera would be suitable provided that filters are used to enhance defects and eliminate extraneous information. If some varieties of date require processing in more than one colour band, the use of filters, reflecting mirrors, or other techniques can still be used to get separate colour band images in real time (colourkeying).

In a practical situation it would be necessary to take several images of the fruit since defects may only show on one face of the date. This may involve two or three cameras taking images of the date simultaneously from different directions. For commercial implementation speed of operation is of paramount importance to keep costs down. In any subsequent work, an investigation of real-time speed of the algorithm is needed.

CONCLUSIONS

1. Illumination is critical. For best results use a diffuse light source to minimize glare, and provide a non-reflecting background.
2. Simple thresholding based on the ratio of defect/good spectral characteristics of dates appears to offer a promising means of detecting several types of defects.
3. Spectrophotometric data can be used directly to obtain both the thresholding parameters used for sorting and the wavelength range in which defect/good date contrast is maximized.

Table I. Description of fruit processed for defects

<table>
<thead>
<tr>
<th>Variety</th>
<th>Date No.</th>
<th>Description</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>'Majhul'</td>
<td>(1)</td>
<td>Good</td>
<td>good</td>
</tr>
<tr>
<td></td>
<td>(2)</td>
<td>Dry end</td>
<td>defect</td>
</tr>
<tr>
<td></td>
<td>(3)</td>
<td>Blisters</td>
<td>defect</td>
</tr>
<tr>
<td>'Zahidi'</td>
<td>(4)</td>
<td>Good flesh, a little &quot;blistered&quot;</td>
<td>good</td>
</tr>
<tr>
<td></td>
<td>(5)</td>
<td>Insect damage</td>
<td>defect</td>
</tr>
<tr>
<td></td>
<td>(6)</td>
<td>Dry, shriveled</td>
<td>defect</td>
</tr>
<tr>
<td></td>
<td>(7)</td>
<td>Wrinkled, red tinge</td>
<td>defect</td>
</tr>
<tr>
<td></td>
<td>(8)</td>
<td>Bird damage</td>
<td>defect</td>
</tr>
<tr>
<td></td>
<td>(9)</td>
<td>Blistered, drying</td>
<td>defect</td>
</tr>
<tr>
<td></td>
<td>(10)</td>
<td>Decay</td>
<td>defect</td>
</tr>
</tbody>
</table>

CANADIAN AGRICULTURAL ENGINEERING Vol. 35, No. 2, April/May/June 1995 137
Fig. 6. Images of 'Majhul' date with dry end: (a) Digitized red band image, \( u_0(m,n) \), taken with optical filters 301A and 22; (b) Mask \( u(m,n) \); (c) Masked date image \( v(m,n) \); (d) Defective pixels image \( w(m,n) \) resulting from analysis of image taken with IR cutoff filter 301A only; (e) Defective pixels image resulting from analysis of image taken with IR cutoff filter 301A and orange filter 22.

Fig. 7. (i) Masked green band date images \( v(m,n) \), and (ii) Defective pixel images \( w(m,n) \) for several of the 'Zahidi' date images (IR cutoff filter only): (a) Good; (b) Dry, shriveled; (c) Decay; (d) Red tinge, wrinkled.
### Table II. Threshold parameters

<table>
<thead>
<tr>
<th>Variety</th>
<th>Optical filter</th>
<th>Wavelength range (nm)</th>
<th>Ave. good grey level</th>
<th>( t_0 )</th>
<th>( S_1 )</th>
<th>( t_1 )</th>
<th>( S_2 )</th>
<th>( t_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>'Majhi'</td>
<td>301A†</td>
<td>520-680</td>
<td>22.07</td>
<td>70</td>
<td>-</td>
<td>-</td>
<td>2.10</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>301A, 22‡</td>
<td>580-680</td>
<td>25.83</td>
<td>64</td>
<td>-</td>
<td>-</td>
<td>1.97</td>
<td>51</td>
</tr>
<tr>
<td>'Zhadi'</td>
<td>301A</td>
<td>450-600</td>
<td>24.63</td>
<td>65</td>
<td>0.16</td>
<td>4</td>
<td>2.00</td>
<td>49</td>
</tr>
</tbody>
</table>

† Infrared cutoff filter  
‡ Deep orange filter

### Table III. Results of image processing algorithm

<table>
<thead>
<tr>
<th>Date No.</th>
<th>Number of pixels</th>
<th>Average grey level</th>
<th>Standard deviation</th>
<th>Pixel value</th>
<th>No. defective</th>
<th>% DEF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Min</td>
<td>Max</td>
<td>&lt;( t_1 )</td>
</tr>
<tr>
<td>1</td>
<td>46,914</td>
<td>22.07</td>
<td>7.50</td>
<td>1</td>
<td>82</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>46,915</td>
<td>32.60</td>
<td>13.33</td>
<td>2</td>
<td>83</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>46,146</td>
<td>24.73</td>
<td>12.56</td>
<td>0</td>
<td>92</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>28,961</td>
<td>24.63</td>
<td>14.98</td>
<td>0</td>
<td>125</td>
<td>17</td>
</tr>
<tr>
<td>5</td>
<td>29,089</td>
<td>23.90</td>
<td>16.20</td>
<td>0</td>
<td>109</td>
<td>420</td>
</tr>
<tr>
<td>6</td>
<td>18,465</td>
<td>38.81</td>
<td>8.51</td>
<td>0</td>
<td>63</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>28,321</td>
<td>19.34</td>
<td>8.31</td>
<td>0</td>
<td>55</td>
<td>28</td>
</tr>
<tr>
<td>8</td>
<td>22,177</td>
<td>42.12</td>
<td>18.05</td>
<td>0</td>
<td>113</td>
<td>21</td>
</tr>
<tr>
<td>9</td>
<td>24,737</td>
<td>31.26</td>
<td>15.38</td>
<td>0</td>
<td>106</td>
<td>12</td>
</tr>
<tr>
<td>10</td>
<td>20,129</td>
<td>19.54</td>
<td>13.79</td>
<td>0</td>
<td>109</td>
<td>1603</td>
</tr>
</tbody>
</table>

† Optical filter 301A  
‡ Filters 301A and 22

4. The algorithm developed has potential as a rapid means of detecting defective dates.

5. The use of optical filters to enhance defects by attenuating light outside the range of interest increases the accuracy of this technique.

6. To validate the results, it is necessary to obtain data for larger samples of each type of defect. Then, much larger samples of dates need to be processed to allow estimation of the sorting errors that would be expected with this technique.

### REFERENCES


