Interactive Exploration of Spatially Distributed Near Infrared Reflectance Data

Hayden Munro [1,2]
[1.] Department of
Information Science,
University of Otago,
P.O.Box 56,
Dunedin, NZ.

Kevin Novins
[2.] Department of
Computer Science,
University of Otago,
P.O.Box 56,
Dunedin, NZ.

George Benwell
Department of
Information Science,
University of Otago,
P.O.Box 56,
Dunedin, NZ.

Alistair Mowat
HortResearch,
Ruakura Research Centre,
Private Bag 3123, Hamilton,
NZ.

Abstract

The display of a vector of data values at a set of spatially-distributed sample points presents some interesting visualisation problems. Typical display devices provide only two spatial di mensions plus colour, making it necessary to design new methods for representing the data. This paper describes a tool under development that allows users to visualise the spatial ripening characteristics of fruit. Sugar, acid and moisture content can be measured using non-destructive Near Infrared Reflectance (NIR) analysis techniques. We introduce the notion of spectrum and spatial tools and show how they may be combined to form a flexible visualisation environment for exploring NIR data. These notions may be generalised to areas such as LANDSAT imagery. We expect that high performance computing systems will enable us to extend our tools so that they can operate at a wide range of spatial scales.

1. Introduction

Scientific visualisation can be described in two ways: as a tool for discovering and under standing, and as a tool for communicating and teaching (DeFanti, 1990). It is used to present information to users in visual forms that appeal to their intuitive understanding. Thus, visualisation tools facilitate the extraction of knowledge from complex datasets. This paper describes a tool that is being developed to allow users to visualise data about the ripening characteristics of fruit. These characteristics, such as sugar, acid and moisture content can be measured using non-destructive

Near Infrared Reflectance (NIR) analysis techniques (Ciurczak, 1995, Murakami, 1993, Murakami, 1992). By comparing the information gained by NIR analysis with the physio chemical properties of the fruit it is possible to relate specific NIR spectral features to desirable product attributes. Once this has been done NIR analysis will be able to be used as an objec tive measure of fruit quality. This information could then be used to manage fruit development and storage processes to maximize market acceptance. It may also be possible to predict fruit maturity and post-harvest characteristics prior to harvest.

The multidimensional nature of the NIR data introduces some interesting visualisation prob lems. The data is four dimensional, whereas the display device only provides two dimensions. It is therefore necessary to discover two dimensional methods for representing the data (McCormick, 1987). Also, the users of the application wish to explore the dataset to discover new features, so it is important to create an application with a high degree of interaction to assist them. In order to tackle these problems this implementation provides a suite of interactive visualisation tools rather than a single display mode. In addition users are given the power to modify the display for their own purposes.

The approaches described here are applicable to all visualisation systems that work with a set of scalar values measured at spatially-distributed sample points. For example, the reflectance data gathered from a fruit using an

NIR probe is exactly analagous to spectral reflectance data measured by a satellite orbiting the earth.

The current tool is useful for exploring spatial relationships in the NIR data across individual fruit. Ideally, we would like to analyse spatial relationships at a wide range of scales, for example comparing different fruit from the same tree, different trees in the same orchard, or different orchards in the same geographic region. Data exploration of this kind is not practical using our current workstation-based system. It will require the high RAM capacity and rapid paging abilities of high performance computing systems. The work described here is part of a continuing joint research between the University of Otago and the Horticulture and Food Research Institute of New Zealand.

2. Data collection

The data for this project is provided by HortResearch. It consists of diffuse reflectance spectra, comprised of individual light intensities measured at 0.5nm intervals, collected from specific locations on the sub-surface of intact fruit. The spatial coordinates of these locations were also provided. So far, datasets have been collected from Gala apples and kiwifruit.

The spatial coordinates were obtained using a Polhemus FasTrak T M device (Smith, 1995). Each fruit was placed inside a cylinder with the major axis perpendicular to the cylinder's straight edge, as shown in Figure 1. A longitudinal line of nine holes had been drilled at 15 degree intervals around the cylinder's circumference. The stylus of the tracker is pushed through each of the holes to measure a position on the fruit surface. These point locations correspond to one eighth of the fruit. The fruit is then rotated 45 degrees to collect the next set of points, and the process is repeated until the whole fruit surface has been sampled. In total 72 points, distributed across eight points of latitude and nine points of longitude, are selected for spectral analysis. This collection method results in a 3D dataset in cylindrical coordinates. These must be converted to Cartesian coordinates before visualisation.

After the FasTrak(TM) device is removed from the guide holes, a Near Infrared Reflectance (NIR) probe is inserted to record the diffuse visible-to-NIR spectrum at the fruit's subsurface at each of the 72 points. The diffuse visible-to-NIR spectrum was sampled at 1100 intervals in the spectral range of 508nm to 1026nm using an Ocean Optics miniature fibre optic probe and charged couple device (CCD) spectrophotometer. The sampled spectrum shows the amount of energy that is reflected at each different measured wavelength (Hall, 1988). There is a minor error term between intensity measurements within a spectral reading. This error is related to the noise characteristics of the 1100 elements contained in the CCD array sensor of the spectrometer.

Commercially available spectral analysis packages such as Grams/32 (Glactic Industries Corp., Salem, NH, USA) may be used to compare reflectance spectra for different points on the fruit. These tools are not sufficient for our task because they do not preserve the spatial relationships between the points that are being compared.

3. Visualisation methods

The data described in the previous section can be considered four dimensional: the points on the fruit's subsurface where wavelengths were measured must be described in 3D in order to maintain their spatial relationships. The intensities measured over the sampled spectrum are the fourth dimension.

Since the data has more dimensions than the display, no single visualisation method will be able to display all aspects of the dataset simultaneously. For this application, the user cannot determine in advance which view will be best. Fortunately the computer's flexibility allows the development of a suite of visualisation tools that present different conceptual views of the data. By using the tools interactively and in concert the user can explore and discover features of interest.

The tools currently provided are divided into two categories: Spectrum tools and spatial tools. Spectrum tools display the intensity at all measured wavelengths for a single point on the fruit surface. Spatial tools display 3D fruit
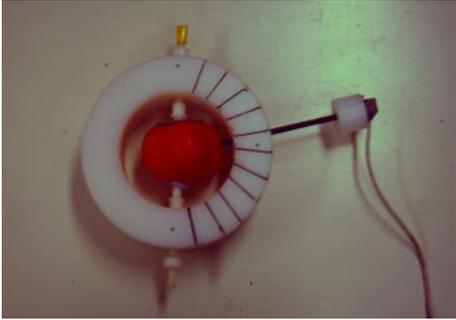
Figure 1. Selecting spatial points on a fruit using a Polhemus tracker.

geometry and intensity values for a single wavelength at all surface points. These categories are described in detail in the following subsections.

The tools were developed using the OpenGL library and the Tcl/Tk scripting language. OpenGL is a standard graphics library designed for real-time applications (Neider, 1993). It is designed to encourage the development of portable programs. OpenGL is a library of functions that can be called from a high-level language, in this instance C is used. OpenGL makes it particularly easy to create and interact with 3D polygonal models.

Tcl/Tk was also chosen for its flexibility. Tk provides a package of tools that are used for creating application interfaces. Tcl is an interpreter that builds the program's interface from command scripts (Ousterhout, 1994, Welch,

1995). Using scripts rather than library calls allows the interface to be modified while the application is running. This feature creates a small decrease in terms of speed, however this is not significant in our application.

The software tools described here are available for many platforms; for the implementations described here, a Silicon Graphics Indigo workstation was used.

3.1 Spectrum tools

Spectrum tools display the intensity at all wavelengths for a single point on the fruit surface. This is essentially a 2D graphing problem: For each of 1100 wavelength samples, we have a single intensity value. We can present the data as a 2D scatterplot, a line drawing, or a histogram. These are presented in Figure 3. Because of its high contrast, we have found the spectral histogram tool to be the most useful.

3.2 Spatial tools

The following views have a common trait that sets them aside from the spectrum tools: Each spatial tool shows multiple points on the fruit surface, but intensity at only one wavelength. This enables comparison of intensities between points, and provides an opportunity to observe their spatial relationships.

3.2.1. 3D Scatterplot View.

The spatial coordinate data can be plotted in a virtual 3D space, as shown in Figure 4. A perspective projection transforms the 3D model into a 2D picture for display (Foley, 1990). The user can interact with the scene using a virtual trackball (Glassner, 1990). The mouse manipulates the viewing position so the model can be seen from any angle, giving the impression of a three dimensional object.

Plotting the points in 3D is not sufficient; it is necessary to find a way to show the spectral data. This can be achieved by displaying the intensities at a single wavelength as colour information. By mapping the largest intensity
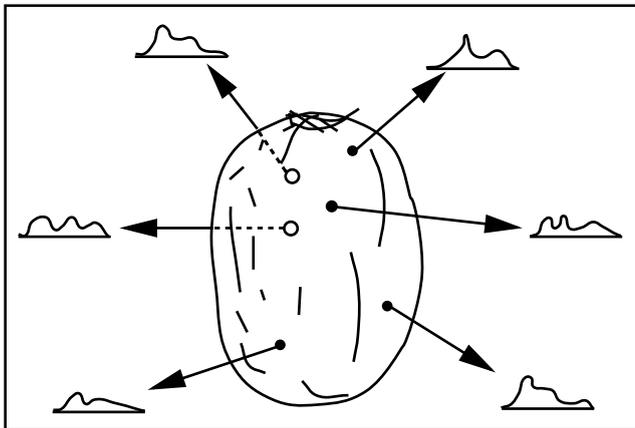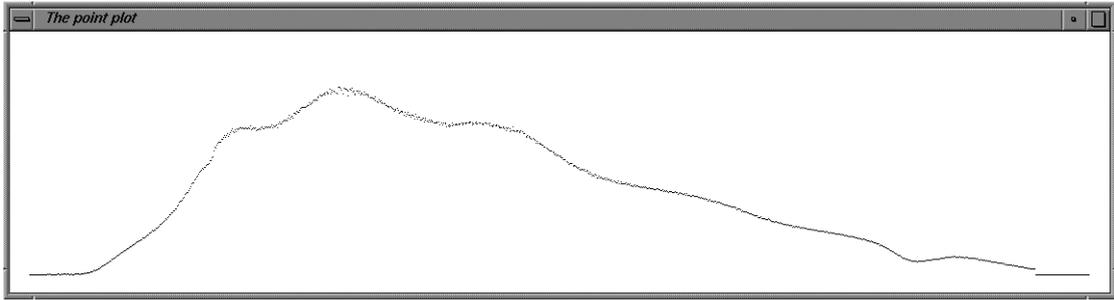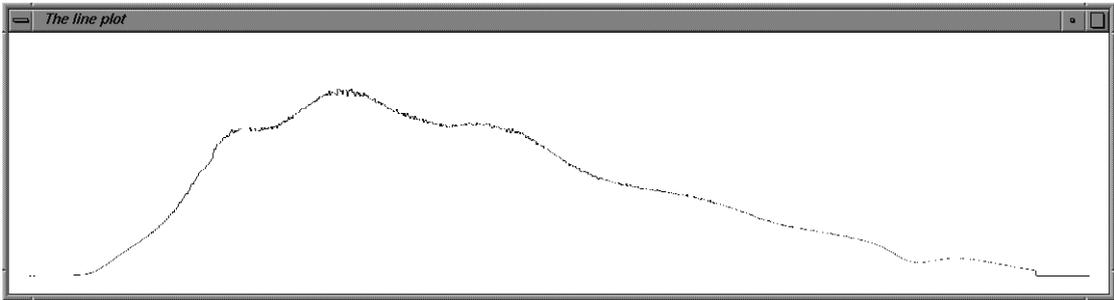

Figure 2. Data is gathered across the NIR spectrum for each sample point, so wavelength becomes a 4th dimension to visualise.
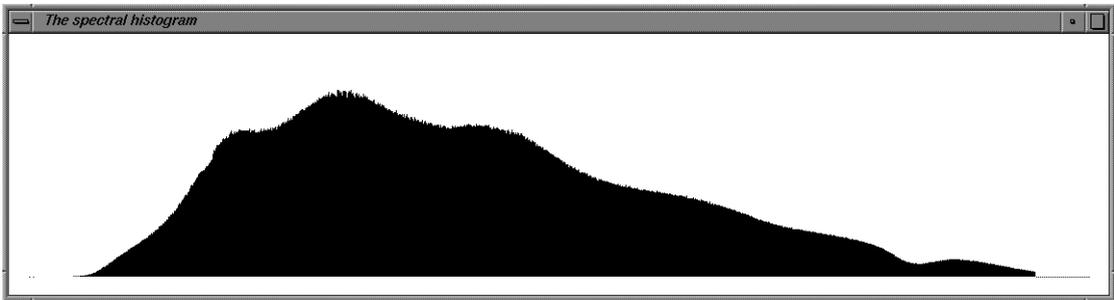
(a)

(b)

(c)

Figure 3. The spectrum tools (a) the point plot, (b) the line plot, and (c) the spectral histogram.

to white and the smallest to black, these and the intermediate values can be shown as levels of gray on the fruit model. A slider can be used to select the individual wavelength to display.

Representing intensity with grayscale has a drawback. If the model is shown on a 24-bit monitor, there is a restriction of 256 grey values. This means that it is only possible to distinguish between 256 intensity values in the dataset. By using the red, green, and blue channels independently it is possible to increase this number by a factor of six. Ware (Ware, 1988) suggests that the lack of colour resolution is

a minor problem when compared with systematic errors that can arise from interpretive effects of the human visual system, such as simultaneous contrast. With this in mind it is more important to reduce these perceptual effects than it is to reduce quantisation of the displayed data. In accordance with this philosophy an approximation to the visual spectrum is made available as a colour map for the wavelength data. The 3D scatterplot view with the visual spectrum colour map is shown in Figures 5a and 5b.

3.2.2. 3D Model View.

The points chosen for each fruit represent a rather sparse

sample. In order for the fruit to be displayed as a "skin" the intensity information is interpolated between points. The method used to interpolate this colour information is known as Gouraud shading (Foley, 1990). Gouraud shading is a popular technique because it is simple and can be computed using graphics hardware. It is useful to extend these points to form a skin for two reasons. First, the shape of the model is clearer if it is represented as a skin. Second, the point colour is interpolated across an area, making it easier to see. The skin is made from polygonal surfaces constructed using adjacency information extracted from the latitudes and longitudes of each point. The 3D scatterplot view is compared to the 3D model view in Figure 5.

3.2.3. The Map View.

The 3D tools are useful for viewing the fruit in a virtual environment but it is restrictive since at any time the back part of the fruit is occluded, leaving the data only partially visible. It is possible to "unwrap" the model view so that the full surface of the fruit is shown in one picture. This is shown in the map view. An example map view is shown in Figure 6. The data collection method provides a natural latitude-longitude coordinate system for the data. In principle, this partitioning allows us to use any of the techniques developed by cartographers for mapping the earth. In order to "flatten" the fruit surface our current tool simply plots longitude as the x-axis and latitude as the y-axis in what is known as an equirectangular projection (Snyder, 1993). All 2D maps of curved objects will contain distortions, and while accuracy of the equirectangular plot is high at the fruit's equator it decreases rapidly as the poles are approached. In this application the effects of distortion are reduced somewhat since the poles of the fruit are not sampled.

3.2.4. The Height Field View.

The map view reduces visual complexity and allows us to use the 3D graphics hardware for another purpose. In the height field view intensities are interpreted as altitudes, and the data is displayed as a 3D relief map. This artificial terrain helps to clarify the relative distance between in-
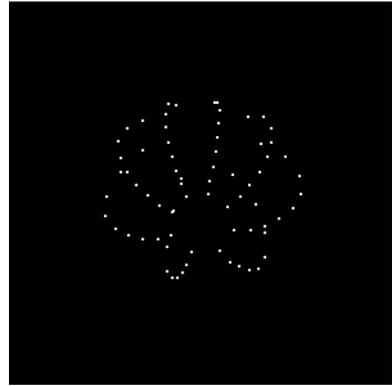

Figure 4. The 3D spatial points.

tensity values, which is difficult to determine when only colour information is used. The height field can be viewed from any angle by manipulating a virtual trackball. The heights also respond interactively to movement of the wavelength slider. It is important to incorporate a stationary reference for this view; otherwise the surface appears as if it is floating in space. The stationary reference is created by vertical lines below the height field in Figure 7.

A new slider is introduced for the height field view. It controls the height of the maximum value that is in the dataset, acting as a scaling constant for the virtual altitudes. This is provided so the user can exaggerate the disparity between similar intensity values that are displayed. It is important to note that applying heights to the intensity values in the 3D model view would not be beneficial. The 3D object is too complex to act as an adequate reference for the changing amplitudes. It would be difficult to distinguish between bumps that represent fruit geometry and bumps that represent intensity data.

4. View integration

The views described in the previous section are combined to create a single application. Each view presents different aspects of the data, so they are made to occupy separate windows that are simultaneously visible. To support exploration using multiple windows, the parameters from each view are linked.

Each point shown on the map view can be selected by clicking or dragging the mouse in the display window. The
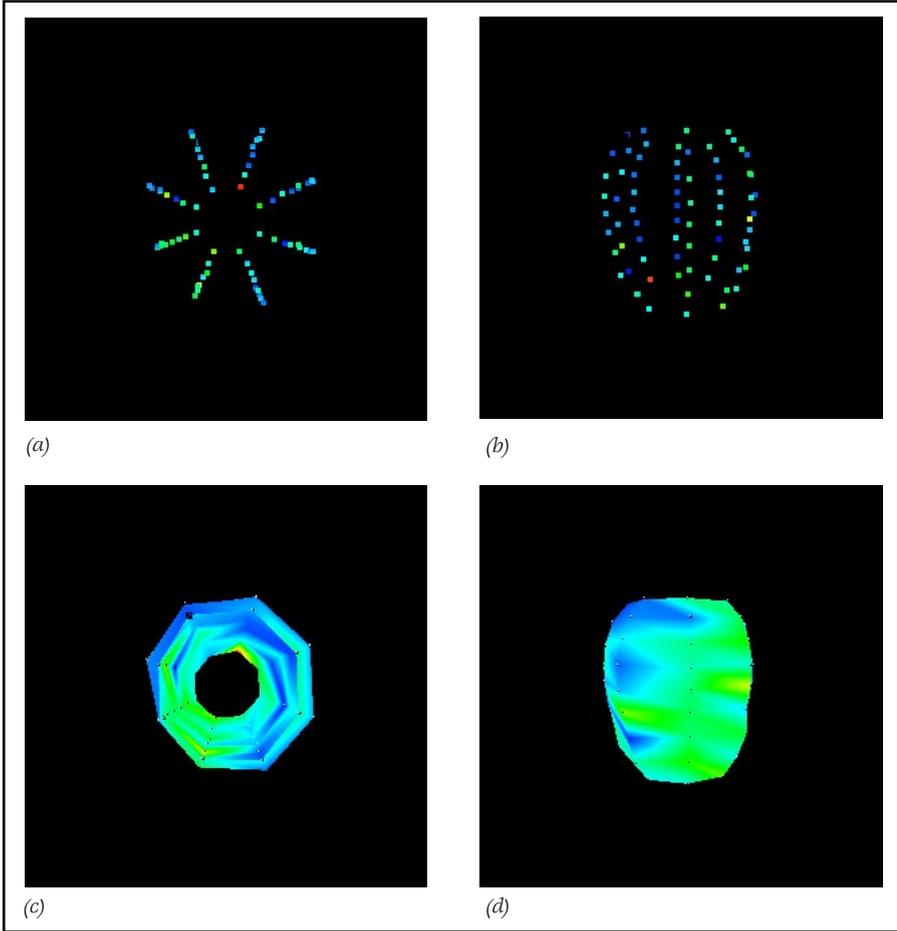
Figure 5. Comparing 3D scatterplot and 3D model views. 3D Scatterplot views for the 737nm response of a kiwifruit are shown from two different angles in (a) and (b). In (c) and (d), the corresponding 3D model views are shown.

closest point to the mouse is highlighted and the spectrum data relating to that point is used to create a spectral histogram. As the mouse is dragged between points the histogram changes interactively. It is possible to select and drag a wavelength indicator on the histogram. As this is done, the wavelength shown in the map view is updated accordingly. This single environment, illustrated in Figure 8, is maintained to avoid confusing the user.

In the current application, the 3D model view and the height field view are also linked to the wavelength selected in the spectral histogram view. A screen shot of the application along with descriptions of the interaction methods is given

in Figure 9. The cohesion between the views provides a fast and effective way to view the data.

5. Conclusion and future work

The integrated application was recently presented to a group of scientists at HortResearch. Their response illustrated that the views create a visualisation environment that encourages exploration of the dataset. Interacting with the raw data graphically gives the user a rapid understanding of the nature of the data. The provision of multiple views has created a great level of flexibility for exploring the dataset. The presentation stimulated many ideas about
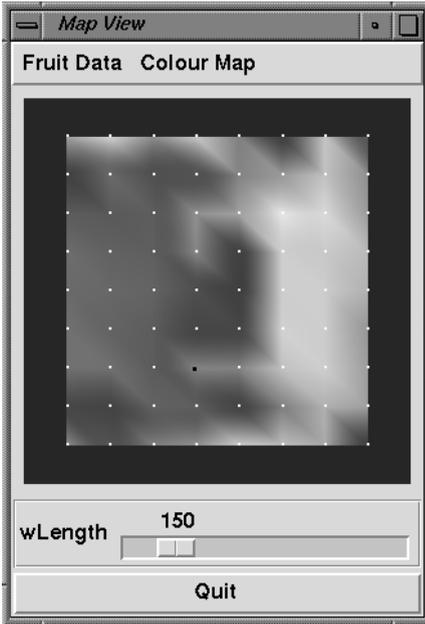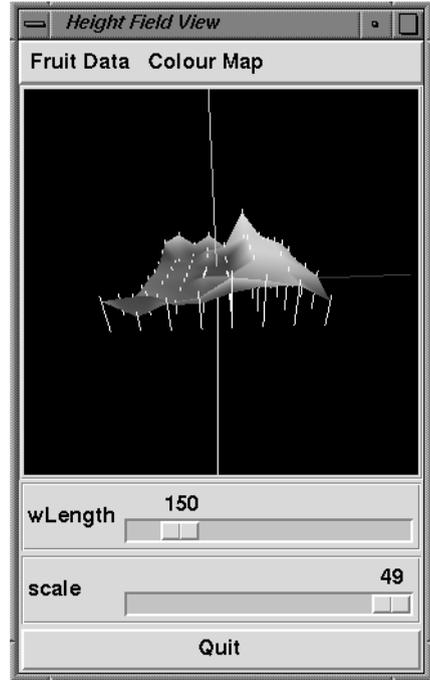
Figure 6. The map view.

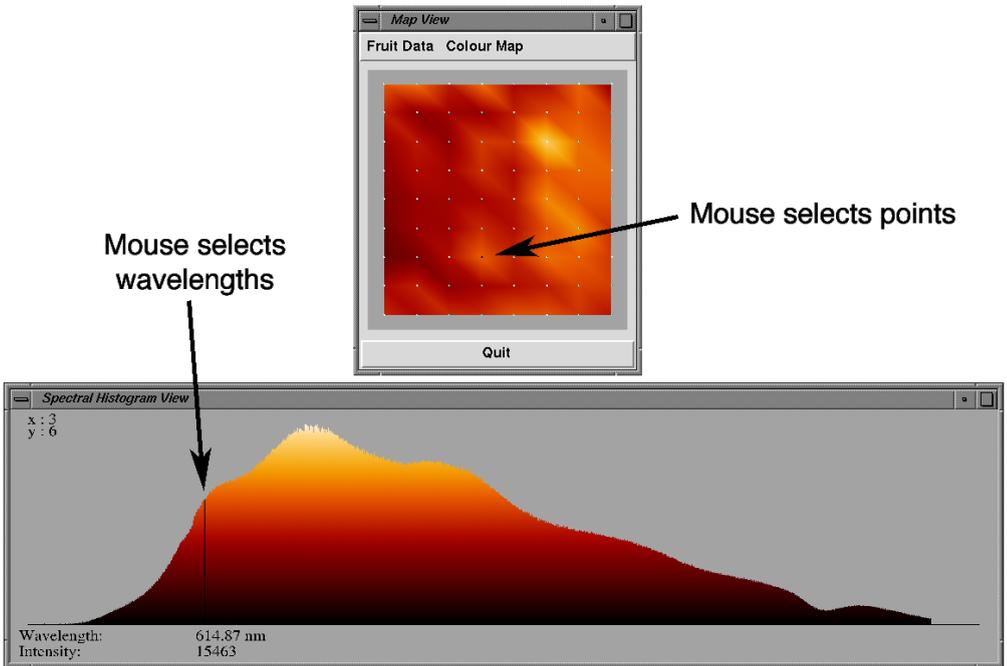*Figure 7. The height field view.*

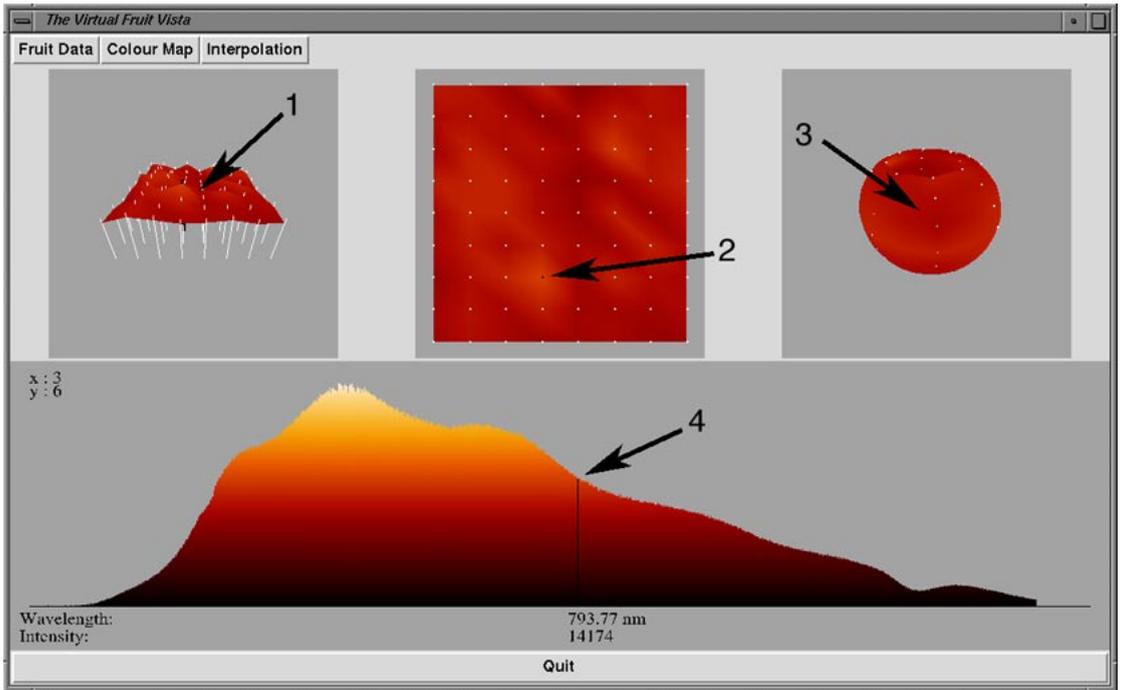*Figure 8. The interaction methods for map and spectral histogram views.*

Figure 9. The integrated application with four interaction methods. In the height field view (1) and the 3D model view (3), the mouse acts as a virtual trackball to rotate the object. Fruit sample points are selected by the mouse in the map view (2), and the selected wavelength is changed by dragging the indicator in the spectral histogram view (4).

how the tools can be enhanced and has defined a clear path for their future development. The tools are considered promising and the project is continuing.

We would like to tune some of the interaction paradigms that are used in the system. For example, it would be useful to be able to select sample points in each of the spatial displays, not just the map view. It would also be helpful if the spatial tools automatically centered the currently selected sample point. This could be done by scrolling the map and height field views and rotating the model view.

Our experience indicates that users gain accurate information about NIR datasets using the visualisation tool. Controlled perceptual studies are necessary to ensure that the user's impres sion of the visualised data matches the reality of the raw data. Possible areas of inaccuracy include the use of a quantised colour scale to display intensity, linear interpolation of intensity between sample points, and the equirectangular map projection.

Because NIR data collection is inexpensive and non-invasive, it is possible to measure the response of an individual fruit throughout its maturation process without removing it from the vine. Portable scanners are already under development (Martinsen, 1996). The most exciting follow up re search will involve the display of NIR data at varying spatial scales and across the temporal dimension. Analysis of this type of data may enable researchers to predict post-harvest quality prior to harvest. They may also be able to provide farmers with advice on how to situate trees and orchards to maximise high-quality yield.

The NIR data for a single fruit at a single point in time requires about 317 kilobytes of storage. When this number is multiplied by the number of apples in a tree, the number of trees in an orchard, the number of orchards in a region, and the number of days to maturity, the memory requirements quickly reach the terrabyte range. In order explore such a dataset interactively, we will need very large RAM

spaces, fast paging capabilities, and high-speed graphics hardware capable of implementing a zoomable interface (Perlin, 1993).

6. Acknowledgements

The software products used in this implementation are Tcl/Tk and OpenGL (a trademark of Silicon Graphics Inc.).

References

E. W. Ciurczak (1995). Use of near-infrared spectroscopy in cereal products. Food testing and analysis, 1(2).

T. A. DeFanti, M. D. Brown, and B. H. McCormick (1990). Visualisation, expanding scientific and en gineering research opportunities. In G. M. Nielson, B. Shriver, and L. J. Rosenblum, editors, Visualisation in Scientific Computing, pages 32-47. IEEE Computer Society Press.

J. D. Foley, A. van Dam, S. K. Feiner, and J. F. Hughes (1990). Computer Graphics, Principles and Practice. Addison Wesley, 2nd edition.

A. S. Glassner (1990). Graphics Gems, chapter 8, pages 462-463. Academic Press.

R. Hall (1988). Illumination and Color in Computer Generated Imagery. Springer-Verlag.

P. Martinsen and P. Schaare (1996). A near infrared imaging spectrometer. Image and Vision Computing New Zealand, pages 109-114, August.

B. H. McCormick (1987). Computer graphics: Special issue on visualisation in scientific computing. Technical Report 21(6), ACM SIGGRAPH, November.

M. Murakami, J. Himoto, and K. Itoh (1993). Analysis of apple quality by near infrared reflectance spectroscopy. Journal of faculty of agriculture, hokkaido university, September.

M. Murakami, J. Himoto, M. Natsuga, and K. Itoh (1992). Analysis of pumpkin quality by near infrared reflectance spectroscopy. Journal of faculty of agriculture, hokkaido university, 65(4):359-366.

J. Neider, T. Davis, and M. Woo (1993). OpenGL Programming Guide. Addison Wesley, 1993.

J. K. . Ousterhout (1994). Tcl and the Tk Toolkit. Addison Wesley.

K. Perlin and D. Fox (1993). Pad: An alternative approach to the computer interface. In Computer Graphics Proceedings, Annual Conference Series, pages 57-73.

G. S. Smith and J. P. Curtis (1995). A fast and effective method of measuring tree structure in three dimensions. Technical report, The Horticulture and Food Research Institute of New Zealand Ltd.

J. P. Snyder (1993). Flattening the Earth. University of Chicago Press.

C. Ware (1988). Color sequences for univariate maps: Theory, experiments, and principles. IEEE Computer Graphics and Applications, 8(5):41-49, September.

B. Welch (1995). Practical programming in tcl and tk. Draft Manuscript.