Using computer vision technology to evaluate the meat tenderness of grazing beef

Y. Q. Tian, D. G. McCall, W. Dripps, Q. Yu and P. Gong

Raw meat surface features from non-grazing animals are reported to be correlated with meat tenderness. However, meat from grazing beef may have different tenderness to that of non-grazing beef due to differences in activity and diet. The feasibility of using meat surface characteristics from grazing beef in New Zealand to estimate meat sensory tenderness was tested. Results from striploin samples from 50 carcasses demonstrated that geometric, spectral and textural characteristics of meat from grazing beef were correlated to meat tenderness assessed by trained tasting panels. Correlations were obtained using a neural network approach (adjusted $R^2 = 0.62$) and a linear multivariable regression technique (adjusted $R^2 = 0.58$).

The red meat industry is increasingly concerned about the quality, as opposed to the increased quantity, of meat products. Several studies have shown that tenderness is the most important criterion for dictating meat eating quality, as judged by consumer taste panels (Savell & others 1989). Currently, carcass grading programs are used to assess carcass quality. The programs are intended to provide some indication of a meat's relative palatability, but these programs have their disadvantages such as subjectivity and inconsistency. A survey of supermarket beef quality conducted in New Zealand (NZ) showed that many grading programs are not accomplishing their objective of ensuring consumer satisfaction (M Paine pers comm). Beef quality assessed from human graders does not always have a strong correlation with taste panel measures. Although some objective mechanisms to measure meat tenderness are available (for example, the Warner-Bratzler Shear Device), these evaluation processes are destructive (Jeremiah & Phillips 2000).

Image processing technologies have promise for assessing meat quality objectively and effectively (Gao & Tan 1996, Kim & others 1998, Vote & others 2003). Previous efforts using image processing techniques for meat assessment were aimed at automating the meat grading process, based on eating criteria (eg palatability, tenderness, flavour, and juiciness). Several recent studies reported that raw meat surface features are potentially good indicators of meat tenderness (Park & others 1998, Li & others 1999, Tan & others 1999, Tan 2004). These findings were based on results from examining meat of grain-fed beef, which is fundamentally different to meat from grazing beef, which typically has less fat, little marbling and more collagen.

The objective of this study was to examine if meat surface features from grazing beef reflect meat tenderness, as has been found for grain-fed beef (Park & others 1998, Li & others 1999). Based on our observations, grazing beef, relative to grain-fed beef, has different muscle size and shape, a greater proportion of collagen or intramuscular fat, and obscure surface texture patterns; as such, we developed a new algorithm for extracting meat surface features for meat from grazing beef. Differences of surface texture (longissimus muscle area, marbling scores and lean color) between beef muscle from grazing and grain-fed cattle have also recently been confirmed by Baublits & others (2004). The meat surface features, extracted using visual technology, were related to tenderness scored by trained sensory panels. The effectiveness of using these surface features for tenderness estimation was evaluated with a linear multivariable analysis and a non-linear neural network method. The tenderness evaluation from the sensory panels was also benchmarked against a mechanical tenderometer evaluation method. In addition, the pH of each striploin sample was analysed to examine the relationship of pH to tenderometer scores. The meat surface features and pH, combined with sensory and tenderometer scores, were used to develop an automated system for consistent and objective meat tenderness assessment.

Materials and methods

Sample collection

Meat samples (*Longissimus dorsi*) from 50 carcasses of NZ grazing beef were collected in two trials. In Trial One, 22 grazing steers from the Whatawhata Research Centre, Hamilton NZ, were killed at a commercial abattoir. Half of these animals had been fed a corn supplement (4 kg/day) in addition to their usual grazing, for eight weeks prior to slaughter to test if the meat features could convey the grain feeding effects on beef tenderness. In Trial Two, 28 grazing heifers were slaughtered at the abattoir and used for sampling. The pH was measured (Orion pH meter) within 24 h of slaughtering of each carcass for both trials, at the ribeye region close to the shoulder end. One portion of striploin (200 mm or 2 kg) was taken from the left side of the carcass. Each portion was vacuum packed, and then frozen 24 h after slaughtering for subsequent imaging, sensory panel processing, and tenderometer testing. The right-side striploins of 22 steers from Trial One were vacuum packed, frozen 24 h after slaughtering, and then sent to

Dr Yong Tian (corresponding author) and Dr Weston Dripps are Assistant Professors, College of Science and Mathematics, University of Massachusetts, Boston, MA 02125, USA; Dr David McCall is Senior Scientist, New Zealand Agricultural Research Institute, Hamilton, New Zealand; Dr Peng Gong is Professor and Qian Yu is a PhD candiate, University of California; Berkeley, USA. Email: yong.tian@umb.edu.

the University of Missouri for parallel sensory evaluation. The processing protocol for the right- and left-side striploins was the same.

The frozen striploins were cut into steaks with a band saw. The steaks were assigned to evaluation treatments as outlined in Figure 1. Cuts #2 and #3 were used for taste panel assessment. Cut #4 was imaged after thawing for 48 h at -1°C and then held for 30 minutes at room temperature. These thawed steaks were imaged in five sessions and were processed in parallel with sensory panel evaluation. Cuts #5–8 were used for tenderometer testing after thawing for 48 h. In order to obtain reliable sensory evaluation of beef tenderness, two sensory panels, one from the USA and one from NZ, evaluated samples from the same animals. The cross-country sensory results were used in verifying the quality of sensory scores.

Sensory protocol

Sensory evaluation was conducted using ten trained panellists, separated to avoid peer influence. The evaluation was performed under red lighting to minimise the effects of visual differences. The sensory scale used by panellists in NZ was a 9-point system (1 = not tender, 9 = very tender). The average score of the ten panellists was used as a measure of the tenderness of each steak sample. Steak samples (20 mm thick) were cooked on broiling pans in electric ovens. A T-type thermocouple was inserted approximately at the centre of each steak to measure the internal temperature. The steaks were first broiled to an internal temperature of 34°C, turned over, and then broiled to a final internal temperature of 68°C. The edges of the cooked steaks were removed and the remaining portion was cut into 1.25 cm cubes. The cubic samples were served in beakers for sensory evaluation. The beakers were preheated in sand baths to 76°C to help maintain the sample temperature.

Sensory protocol and cooking methods used at the University of Missouri were the same as those used in New Zealand except for the scoring scale. The University of Missouri team used an unstructured 16-point line scale (0 = not tender, 16 = very tender) for scoring. The US sensory evaluation system was to test the linear correlation between the two scale systems.

Tenderometer examination

Tenderometers are mechanical devices designed to measure the force needed to displace meat within the sample (Jeremiah & Phillips 2000). The tenderometer contains a set-up of stainless steel needles, sometimes referred to as probes. Tenderness or toughness of the samples is calculated by shear force measurement. The MIRINZ Tenderometer, as described by Macfarlane & Marer (1966) was used to examine the toughness of the cooked beef samples. The beefsteaks used for the tenderometer test were cooked in the same way as those for the sensory panel evaluation. This mechanical approach was used to benchmark the visual imaging technique to the sensory evaluation.

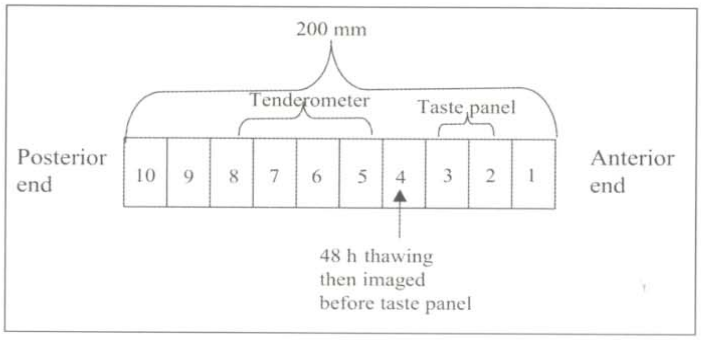

Figure 1. Evaluation treatments for each striploin sample

Image processing and feature extraction

All meat samples were imaged using an image acquisition system that included a colour digital camera (JVC KY-30 3CCD) and a QUINTEK mosaic colour frame grabber card. The lighting box used with the camera was 80 cm in height with a 50 cm^2 floor area. Four Philips high frequency fluorescent tube lamps were mounted on the ceiling. The internal walls of the box were painted white to project uniformly diffused light of the same intensity over the samples.

An algorithm was developed to segment the steak images into various object images. The segmentation algorithm is similar to the one designed by Gao & Tan (1996), but modifications were made to the algorithm to better suit meat from grazing beef. The segmentation was automated. Each image of *Longissimus dorsi* (LD) muscle contains intra-muscular fat (marbling) and lean muscle (Gerrard & others 1996). After the segmentation, the object images of the LD, the marbling, and the ribeye were used for analysis. The LD image contained only the LD muscle (without other tissue or marbling in it). The marbling image had only the marbling inside the LD. The ribeye image was a combination of LD and marbling images, ie the LD with marbling. From these object images, colour, marbling and textural features were extracted.

Features characterising the colour were extracted for the LD muscle only. The mean, standard deviations, and third moments for each of R, G and B (red, green and blue) were calculated individually. The standard deviations indicate the non-uniformity of the colours, and the third moments show the skewness or imbalance of colours.

To measure marbling abundance, statistics on marbling flecks were obtained based on number and area densities for different size categories based on fleck size (A) (such as $A \leq 5$, $5 < A < 50$, $A \geq 50$). The marbling density features computed were:

1. The number of marbling flecks in each size category divided by the ribeye area
2. The number of all marbling flecks divided by the ribeye area
3. The total marbling area for each size category divided by the ribeye area
4. The area of all marbling flecks divided by the total ribeye area.

These four density features are commonly used to describe the marbling abundance (Gerrard & others 1996, Li & others 2001).

Image texture generally refers to geometric properties such as fineness, coarseness, smoothness, granulation, randomness, and graininess of an image. The texture of the beef muscle images reflects muscle fibre characteristics such as size and arrangement and consequently may

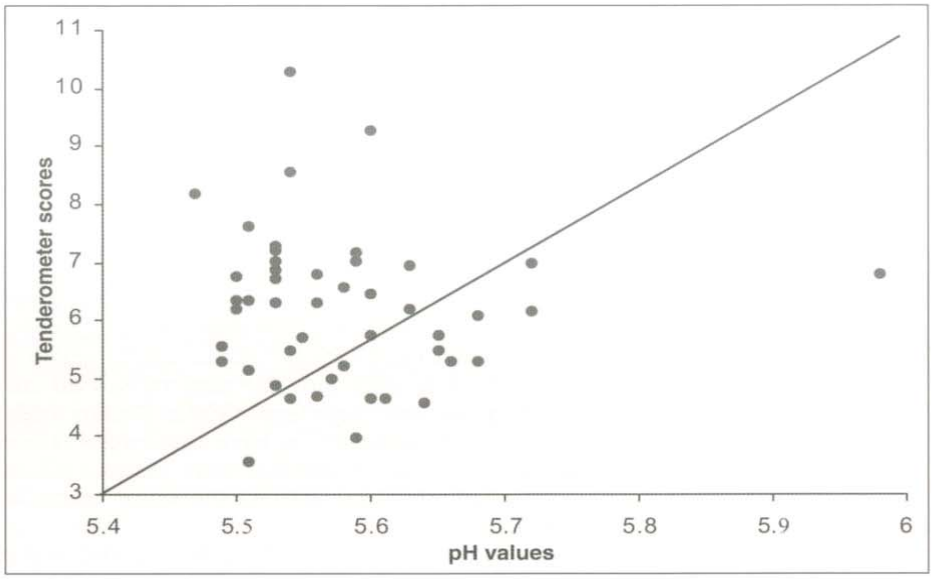

Figure 2. Relationship of pH values to tenderometer scores (kg/cm²)

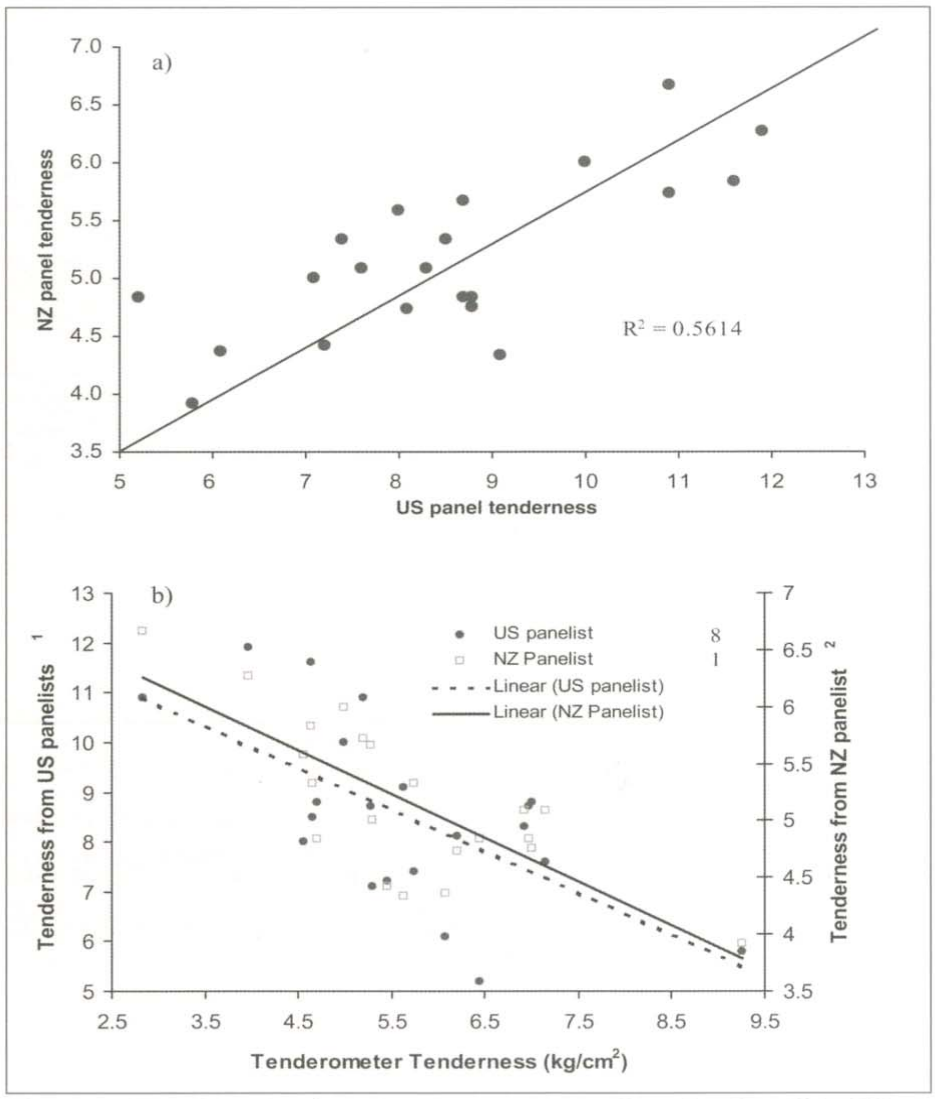

Figure 3. Benchmark of tenderness score produced using NZ v US taste panels (top), and of both panel scores against tenderometer values (bottom).

reflect beef tenderness. Two textural extraction algorithms were adapted for this application: co-occurrence (Haralick & others 1973) and Pixel Value Run Length (Gao & Tan 1996).

The co-occurrence algorithm is used for studying the spatial dependence of pixel values. The spatial dependence of pixel values may be represented by a matrix $P_{d,q}$ with entry $P_{d,q}(i,j)$ being the relative frequency for two pixels d-pixels apart in direction q to have values i and j respectively (Haralick & others 1973). Only direction q = 0 was used, and nine values for d (1 to 9) were tested.

For the Pixel Value Run Length algorithm, a pixel value run is a set of contiguous pixels having the same or similar pixel value in the same direction. A long pixel value run indicates that the pixel values do not change greatly over a long distance, whereas a short run shows significant spatial variability. Pixel value run lengths can thus be used to characterise the spatial variability of pixel values in an image texture. The pixel value run lengths were computed from each image and used to generate a histogram. The run length histogram was defined as P(R, q, T), where P stands for frequency, R is run length in number of pixels, q is the run direction on the image plane, and T is a pixel value band thickness (Gao & Tan 1996). The band thickness is the specified maximum difference in the pixel values included into a run. Nine values from 2 to 10 were tested in this study. Since there was no obvious directionality in the muscle images, only q = 0 (horizontal) was used. The run lengths were computed for each colour function (R, G or B) of a muscle image. From the pixel value run length histograms, the following three features were calculated:

1. Average run length
2. Standard deviation showing variations in run lengths, and
3. Third moment, indicating the skewness or imbalance of run lengths.

Data analysis

A statistical method, best subset, was applied to these extracted meat surface characteristics in order to eliminate insignificant features with

respect to the sensory tenderness scores. The remaining features were used to develop a multivariable linear regression model for predicting sensory scores. A non-linear neural network approach (backward propagation) was also performed to estimate tenderness. The linear regression analysis results were benchmarked against those from the non-linear analysis.

Results and discussion

Effect of grain supplement on beef tenderness

Originally, the experiment was partially designed to test if tenderness scores between grain supplemented and non-grain supplemented beef would be distinguishable. Unfortunately, the statistical analysis showed that the difference in beef tenderness based on the taste panel scores from the two finishing treatments was not significant (non-grain supplement = 5.3 ± 0.8; grain supplement = 4.9 ± 0.7; $p = 0.327$). This result suggests that the grain supplement period may have been too short to influence beef tenderness. Attempts to analyse meat tenderness between the two treatments using computer vision technology were consequently discontinued.

Table 1. The best 20 variables correlated to taste panel's tenderness scores

Variables	Variables	Variables	Variables
B_mean	$d_1\_f_1$	$d_6\_f_{12}$	$rl_{10}\_G\_3$
$d_3\_f_{13}$	$d_3\_f_1$	$rl_7\_R\_3$	$rl_5\_G\_3$
$d_9\_f_3$	$d_5\_f_1$	$rl_5\_R\_2$	$rl_{10}\_G\_2$
$d_2\_f_7$	$d_1\_f_3$	$rl_2\_R\_2$	$rl_8\_G\_3$
$d_1\_f_{13}$	$d_1\_f_8$	$rl_8\_G\_2$	$rl_8\_B\_2$

Where B_mean: the average of blue value;
Labels: $d_x\_f_i$ are variables based on Co-occurrence algorithm, where d_x stands for distance in x; and f_i stands for ith feature given in Haralick & others (1973).
Labels: $rl_{i\_c\_j}m$ stands for characteristics of Run Length: where i: length; and j: j^{th} momentum; C: colour R or G or B

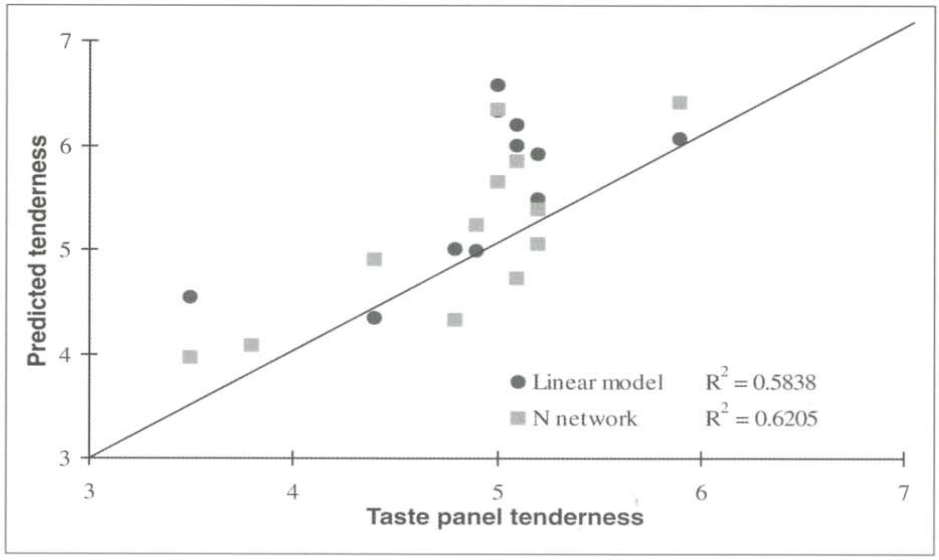

Figure 4. The neural network and linear regression model predictions versus tenderness scores from taste panelists

Usefulness of pH to tenderness estimation

The pH of each striploin sample was plotted against the corresponding scores of tenderometer evaluations (Figure 2). The pH of most samples ranged between 5.4 and 5.8, but tenderometer scores were widely scattered between 3 and 11 kg/cm^2. The lack of correlation between pH and tenderness in this experiment is consistent with results reported by Watanabe & others (1996) and Silva & others (1999), however, they reported that beef tenderness has a curvilinear relationship for pH values between 5.8 and 6.2. We were unable to test this hypothesis since there were only a few samples with a pH higher than 5.8.

Tenderness evaluation

Sensory evaluation is largely subjective and may be biased by cultural influence, human emotions, and eating habit. To ensure the quality of sensory evaluation, meat samples of the same animals in Trial One were appraised in New Zealand and the United States. Sensory teams in both countries were comprised of trained taste panellists. Despite the cultural difference of sensory groups, the benchmark results were consistent and showed good overall correlation ($R^2 = 0.56$) between the tenderness evaluations from both countries (Figure 3a). Aside from the tenderometer's destructive nature, it is a more objective way of measuring meat tenderness. Tenderness scores from the US and NZ sensory panelists on the meat samples in trial one were compared with the tenderometer evaluation separately (Figure 3b). The respective tenderometer correlation with tenderness is $R^2 = 0.575$ for NZ panelists and $R^2 = 0.3482$ for US panelists. The higher correlation from the NZ panellists may indicate that they were better able to discern differences in tenderness than the US panellists.

Empirical predictive model

A stepwise statistical analysis method was used to select potential variables for tenderness assessment from the extracted raw meat surface characteristics. Twenty features were identified as significant in predicting meat tenderness (Table 1). Of these, 19 were textural features: 10 were identified with the co-occurrence algorithm, and nine were identified from the run length algorithm. Only one colour variable was chosen, and no variables were selected from the marbling analysis. The statistical analysis showed that textural features were most significant for tenderness prediction in meat samples from grazing animals.

Meat marbling features in grazing beef did not contribute to tenderness prediction. Very little marbling can be found in NZ beef. To further complicate the analysis, not all white objects within the LD muscle are marbling. Some white objects are connective tissues (collagen). Marbling may improve meat tenderness (McDonald & Chen 1991), but connective tissues decrease meat tenderness (Harris & others 1992). The two features would serve to cancel each other out.

Meat colour had poor correlation

with sensory scores of tenderness based on the statistical analysis. Meat colour is associated with many external factors such as exposure time, lighting, camera settings and temperature of the imaging environment. Changing meat chemical and physical conditions also provide difficulties in controlling any of these colour factors. Previous research identified that colour can be useful for tenderness prediction (Wulf & others 1997), but as an indicator it is less robust as colour changes with time and lighting conditions continuously (Harris & others 2001, Li & others 2001).

The linear regression equation was created for predicting tenderness with the 20 selected variables. The 78 beef samples from both Trials One and Two (left and right carcasses) were used for the analysis. Data from 66 randomly selected samples were used for the creation and calibration of the model, and another 12 samples were used for model verification (Figure 4). The created linear model was able to explain 58% of the variation of the tenderness (adjusted $R^2 = 0.58$).

Neural network model for prediction of tenderness

Meat surface features and sensory evaluations were similarly used for developing and calibrating a neural network model with a backward propagation algorithm for predicting meat tenderness (Kosko 1992). The same 20 variables used for the linear model were extracted from the same 66 meat samples and used for the model development. The learning rate and training goal were set to 0.05 and $1e^{-9}$. Eleven neurons were set in the network structure. The trained neural network was tested against the results for 12 meat samples separate from the training data. The created neural network model was able to explain 62% of the variation of the tenderness (adjusted $R^2 = 0.62$), which is slightly better than the result from the linear regression method ($R^2 = 0.58$). Neural networks consistently have performed better for meat samples than using statistic methods for previous experiments (Li & others 1999). The seemingly marginal advantages of using a neural network approach over a more traditional statistical method are not significant enough to suggest that the relationship between meat surface features and sensory evaluation is non-linear.

Concluding remarks

The results from this study suggest that selected geometric textural characteristics of raw meat of grazing beef can be used to objectively predict meat tenderness, as has been previously shown for non-grazing beef. A neural network approach was able to predict tenderness slightly better than a linear regression approach. Unlike meat of beef from feedlot systems, marbling features extracted from beef LD muscles from grazing beef did not contribute to meat tenderness prediction. The image processing analysis was unable to identify any difference between grazing beef and grazing beef that had received an eight week grain supplement. Several meat scientists advised that the technology could improve its functionality by considering pattern characteristics such as bundles on raw meat surface. Also, separating marbling and connective tissues could improve the model performance of prediction.

Acknowledgements

This research is partially supported by Meat New Zealand. We are grateful to Martin Upsdell for his valuable advice on the statistical analysis.

References

Baublits, RT & others. 2004. Carcass and beef color characteristics of three biological types of cattle grazing cool-season forages supplemented with soyhulls. Meat Sci. 68: 297-303.

Gao, X & Tan, J. 1996. Analysis of expanded-food texture by image processing, Part I: Geometric properties. J Food Process Eng. 19: 425-444.

Gerrard, DE, Gao, X & Tan, J. 1996. Beef marbling and color score determination by image processing. J Food Sci. 61: 145-148.

Haralick, RM, Shanmugam, K & Dinstein, I. 1973. Textural features for image classification. IEEE Trans. on Sys. Man and Cybernetics, SMC-3: 610.

Harris, JJ, Miller, RK & Savell, JW. 1992. Evaluation of the tenderness of beef top sirloin steaks. J Food Sci. 57 (1): 6-9, 15.

Harris, SE & others. 2001. Antioxidant status affects color stability and tenderness of calcium chloride-injected beef. J Animal Sci. 79: 666-677.

Jeremiah, LE & Phillips, DM. 2000. Evaluation of a probe for predicting beef tenderness. Meat Sci. 55: 493-502.

Kim, ND & others. 1998. Ultrasound image texture analysis for characterizing intramuscular fat content of live beef cattle. Ultrasonic Imaging 20: 191-205.

Kosko, B. 1992. Neural Networks and Fuzzy Systems. Prentice Hall, Englewood Cliffs, NJ.

Li, J & others. 1999. Image texture features as indicators of beef tenderness. Meat Sci. 53: 17-22.

Li, J, Tan J & Shatadal, P. 2001. Classification of tough and tender beef by image texture analysis. Meat Sci. 57: 341-346.

Macfarlane, PG & Marer, JM. 1966. An apparatus for determing the tenderness of meat. Food Technol. 20: 134-135.

McDonald, TP & Chen, YR. 1991. Visual characterization of marbling in beef ribeyes and its relationship to taste parameters. Trans. ASAE 34: 2499-2504.

Park, B & others. 1998. Near-infrared reflectance analysis for predicting beef longissimus tenderness. J Animal Sci. 76: 2115-2120.

Savell, JW & others. 1989. National consumer retail beef study: Interaction of trim level price and grade on consumer acceptance of beefsteaks and roasts. J Food Quality 12: 251-274.

Silva, JA, Pararata, L & Martins, C. 1999. Influence of ultimate pH on bovine meat tenderness during ageing. Meat Sci. 52: 453-459.

Tan, JL. 2004. Meat quality evaluation by computer vision. J Food Eng. 61 (1): 27-35.

Vote, DJ & others. 2003. Online prediction of beef tenderness using a computer vision system equipped with a BeefCam module. J Animal Sci. 81 (2): 457-465.

Watanabe, A, Daly, CC & Devine, CE. 1996. The effects of the ultimate pH of meat on tenderness changes during ageing. Meat Sci. 42: 67-78.

Wulf, DM & others. 1997. Using objective measures of muscle color to predict beef longissimus tenderness. J Animal Sci. 75: 684-692. □