

Automatic body condition scoring system for dairy cows based on depth-image analysis

Kaixuan Zhao^{1,2,3*}, Anthony N. Shelley⁴, Daniel L. Lau⁴, Karmella A. Dolecheck⁵, Jeffrey M. Bewley⁶

(1. College of Agricultural Equipment Engineering, Henan University of Science and Technology, Luoyang 471023, China;

2. Key Laboratory of Equipment and Informatization in Environment Controlled Agriculture, Ministry of Agriculture and Rural Affairs, Hangzhou 310058, China;

3. College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China;

4. Department of Electrical Engineering, University of Kentucky, Lexington 40546, USA;

5. Department of Animal and Food Sciences, University of Kentucky, Lexington 40546, USA;

6. Alltech Inc., Nicholasville 40356, USA)

Abstract: Body condition score (BCS) is an important management tool in the modern dairy industry, and one of the basic techniques for animal welfare and precision dairy farming. The objective of this study was to use a vision system to evaluate the fat cover on the back of cows and to automatically determine BCS. A 3D camera was used to capture the depth images of the back of cows twice a day as each cow passed beneath the camera. Through background subtraction, the back area of the cow was extracted from the depth image. The thurl, sacral ligament, hook bone, and pin bone were located via depth image analysis and evaluated by calculating their visibility and curvature, and those four anatomical features were used to measure fatness. A dataset containing 4820 depth images of cows with 7 BCS levels was built, among which 952 images were used as training data. Taking four anatomical features as input and BCS as output, decision tree learning, linear regression, and BP network were calibrated on the training dataset and tested on the entire dataset. On average, the BP network model scored each cow within 0.25 BCS points compared to their manual scores during the study period. The measured values of visibility and curvature used in this study have strong correlations with BCS and can be used to automatically assess BCS with high accuracy. This study demonstrates that the automatic body condition scoring system has the possibility of being more accurate than human scoring.

Keywords: body condition score, depth-image processing, curvature analysis, machine learning, precision dairy farming

DOI: 10.25165/j.ijabe.20201304.5655

Citation: Zhao K X, Shelley A N, Lau D L, Dolecheck K A, Bewley J M. Automatic body condition scoring system for dairy cows based on depth-image analysis. *Int J Agric & Biol Eng*, 2020; 13(4): 45–54.

1 Introduction

The metabolizable energy stored in fat and muscle is vital to maintaining dairy cows. Body weight alone is not a good representative of bodily energy reserves, as the relationship between these variables is affected by parity, stage of lactation, frame size, gestation, and breed^[1-4]. Body condition was defined as the ratio of body fat to nonfat components in the body^[5]. Because direct measurements of body adiposity are difficult and expensive, multiple body condition score (BCS) evaluation systems have been developed to indicate and evaluate the relative amount of subcutaneous body fat or energy reserves of a live cow^[6].

BCS evaluations can be used to determine whether a cow is in

the proper condition for each stage of the lactation cycle. Using BCS information, appropriate dietary changes can be made to maximize the performance of cows^[7]. Cows with unfavorable BCS are at high risk of metabolic and other diseases in the peripartum period. At calving, BCS must be sufficient to allow maximal milk production and health, but excessive BCS at this stage may result in calving difficulties and animal losses^[6]. In early lactation, BCS is a vital indicator of excessive weight loss, which can lead to metabolic disorders and should be avoided^[8]. At dry-off, parturition, and throughout the lactation cycle, BCS evaluations can be used to identify cows that are at risk of milk fever, mastitis, lameness, and infertility^[9]; therefore, BCS is an important management tool for maximizing milk production and reproductive efficiency^[9,10] and even preventing potential disease and lameness^[6,11].

It is estimated that less than 5% of US dairy herd managers regularly assign BCS values to their cows^[12]. Most dairy farms that conduct BCS in the US use a 5-point system^[13]. This system measures the relative amount of subcutaneous fat in 0.25-point increments, where 1 denotes a very thin cow and 5 indicates an excessively fat cow^[14]. The manual scoring of body condition requires experienced personnel with adequate training. Although a well-trained scorer can score one cow in a short amount of time, it is time consuming to score all the cows in a large herd on daily basis. Additionally, the perception of fatness and the

Received date: 2020-01-07 **Accepted date:** 2020-06-08

Biographies: Anthony N. Shelley, PhD, research interests: computer vision technologies, Email: anshel2@uky.edu; Daniel L. Lau, PhD, Professor, research interests: image and signal processing, Email: dllau@uky.edu; Karmella A. Dolecheck, PhD, Associate Professor, research interests: livestock wellness and performance, Email: karmella.dolecheck@uky.edu; Jeffrey M. Bewley, PhD, Senior Adviser, research interests: precision dairy farming technologies, Email: jbewley@Alltech.com.

***Corresponding author:** Kaixuan Zhao, PhD, Lecturer, research interests: computer vision applied in dairy precision farming. College of Agricultural Equipment Engineering, Henan University of Science and Technology, Luoyang, China. Tel: +86-379-64877717, Email: kx.zhao@haust.edu.cn.

understanding of the BCS guidelines vary from person to person, which causes inconsistencies in data from different scorers.

The dairy industry and research community have recognized the need for a quick and inexpensive but accurate technology with which to automatically measure body fat on cows in different stages of lactation. Bewley, et al.^[15] explored the possibility of developing an automatic BCS system based on 2D digital images. A total of 23 points were selected manually from each image to analyze the contour and shape of cows. The study showed that the hook angle, posterior hook angle, and tailhead depression were significant predictors of BCS. According to the above method, Azzaro, et al.^[16] developed an application with which to extract the 23 anatomical points. In the study, a shape descriptor based on principal component analysis was built and tested. Validation testing showed that the average error of the polynomial model was 0.31 from the manual scores. Several other studies have also been conducted to automate BCS evaluation based on 2D and thermal image processing technology^[17-19].

As three-dimensional (3D) images contain information on the depth dimension of the body surface of a cow, they have great potential to improve the accuracy of automatic BCS systems. Weber et al.^[20] developed an automatic 3D optical system to estimate the backfat thickness (BFT) of cows; this system has great potential for scoring body condition. The correlation coefficient between the observed and estimated BFT was 0.96. Fischer, et al.^[21] used a 3D camera to capture the surface of the rear of the cow, and four anatomical landmarks were identified manually from the surface. The principal component analysis was applied to the dataset, and the coordinates of these surfaces in the principal component space were used to build a multiple linear regression model with which to assess BCS. The system still requires some level of human involvement. DeLaval^[22] released an automatic BCS system that provides continuous and daily BCS readings^[23]; however, the high cost of the system hinders the popularization and application of the BCS system.

Spoliansky, et al.^[24] developed an automatic BCS system using a low-cost 3-dimensional Kinect camera. In that study, 14 features were used to build prediction models. When the model was applied, 94% of the errors were under 0.75 BCS points compared to manual scores, and all errors were under 1 BCS point. However, the models required weight information in addition to depth images. Alvarez, et al.^[25] used a depth camera to capture the image of the back of dairy cows. After removing the background, the images directly input into a convolution neural network for training. Finally, 1158 images were used for model training and 503 images were tested. The results show that 78% of the samples with a BCS error less than 0.25.

In light of the above studies, it is apparent that, because of either cost or accuracy, proposed automated BCS systems fall short of the requirements for farm daily use. As such, there is a considerable need for further research to improve the practicality of those systems to a higher level. Therefore, the objective of this study was to (1) develop a fully automatic 3D computer vision-based system with which to assess the BCS of cows with high accuracy ($MAE < 0.25$); (2) explore the possibility of the automated system being more accurate than human scoring. In order to obtain the information directly related to BCS from the depth image, four specific areas (bones or body structure) on the back of the cow were located automatically. Their visibility or curvature was analyzed for quantitative evaluation of fatness in that area. The machine learning algorithms and regression analysis

model were constructed to score the body condition accurately. The stability and consistency of the system were tested and analyzed with the comparison of automated BCS to human scores on daily basis.

2 Materials and methods

2.1 System setup

The data for this study were collected on the University of Kentucky Coldstream Dairy Research Farm. A group of 94 Holstein cows was milked twice a day in the morning (5:00 AM) and afternoon (4:00 PM). All cows left the parlor through a roofed walkway and returned to the free-stall barn after milking. The walkway was paved with a concrete slab floor with walls on both sides. The width of the walkway was 1.03 m, which restricted the movement of the cows.

A PrimeSense™ Carmine 1.08 RGB+depth sensor (PrimeSense™, Tel Aviv, Israel) was used to capture depth images of each cow's back contours as it walked through the return alley. The camera system was placed 3.05 m above the floor of the walkway, with its field of view covering the entire width of the walkway. The camera was connected to a computer in the dairy office via a 30m active repeater USB 2.0 cable. The images taken by the camera were sent to the computer and stored on the hard drive. The resolution of the depth images was 320×240 pixels with a frame rate of 30 fps.

As shown in Figure 1, the walls of the walkway ensured that the cows were always located in the middle area of the frame and oriented roughly parallel to the midline of the image.

To obtain a single file for each cow, we developed software to record the depth images as each cow passed beneath the system. Four fixed lines in the image scene were used to trigger and stop the recording. As the cow walked from the left to the right side of the scene, data recording started when the cow's nose reached the fourth line. Then, depth frames were captured continuously until the tail end of the cow passed the first fixed line. At the beginning of the recording, the software saved the initial image (without a cow in it) as the background image to perform background subtraction later.

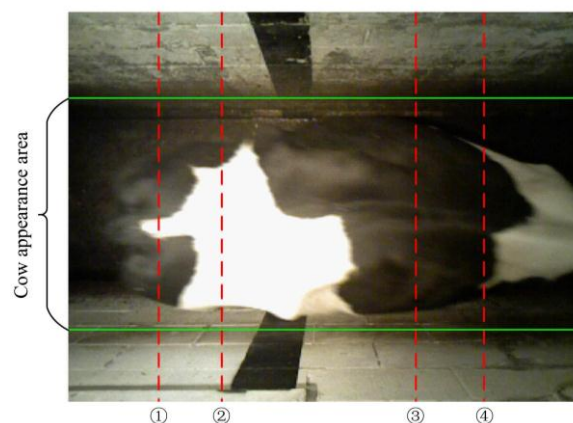


Figure 1 Recording images when four reference lines are reached

2.2 Data acquisition

Depth images of 94 cows were obtained twice a day from April 1st to June 7th, 2014. Over this time, three independent human scorers manually scored every cow in the group once a week on the same day when possible or, at most, within a few days of one another. The median of these three scores from one cow in one week was assigned as the score for the cow that week. Within-cow outliers were removed by comparing the BCS obtained

during successive weeks. When a given BCS differed from preceding and subsequent scores by more than ± 0.25 , the score was removed from the dataset. The objective of this editing technique was to remove individual BCS values that were clearly inconsistent with scores for an individual cow in a short time frame. Because very fat cows were rare in the herd (less than 1.5%), cows with scores above 3.75 were assigned a score of 3.75 to eliminate outliers. Over the two-month data collection period, a total of 94 unique cows were examined at various stages of lactation and levels of body condition. The camera recordings of each cow in a given week were paired with the human-evaluated BCS for that week. The dataset contained 4820 images from 94 cows and their related BCS values from 2.25 to 3.75 (7 classes).

N images were randomly selected from each class to build a training dataset. With M_i denoting the number of images in each class, N is the minimum in M_i . Through this approach, the training model was trained at the same level for each class, and overtraining for the larger classes was avoided. Table 1 illustrates the number of images (M_i) among different BCS classes and the proportion of training data in each class. Eventually, 952 images were selected for training data, i.e. 136 images from each BCS class ($N = 136$). The BCS values of the training dataset were normally distributed. The classification and regression models were calibrated on the training dataset, and tested on training and entire dataset, respectively.

Table 1 Number of images and proportion of training data among BCS classes

BCS	2.25	2.50	2.75	3.00	3.25	3.50	3.75
M_i	467	1414	1662	516	308	317	136*
$(N M_i^{-1})\%$	29.12	9.62	8.18	26.36	44.16	42.90	100

Note: * indicates the number of training images in each class was equal to the minimum among classes ($N=136$), i.e. 136 images were randomly selected from each class to build a training dataset that contains 952 images.

2.3 Feature definition

The image features used in this study were selected because they are potential indicators of BCS. Body fat reserves on the sacral ligament were measured using the convex hull. Surface curvature was used to define the sharpness of the hook and pin bones. Overall back fatness was also evaluated.

Convex hull. As shown in Figure 2, the sacral ligament has a concave curve on thin cows, and the curve is less concave on fat cows. As a thin cow gains fat on the sacral ligament, the concave parts of the curve are filled in, and the concave curve finally becomes a convex curve. In this paper, the convex hull is defined as a convex curve that is tangent to the concave line and lies at a minimum distance from it; therefore, the convex hull can be a tool with which to simulate how the sacral ligament would look if a cow gained body fat. In this study, the visibility of the sacral ligament was measured by the space between the convex hull and the outline of the sacral ligament.

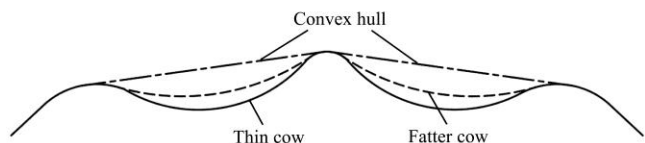


Figure 2 Simulation of the sacral ligaments of a thin cow and a fat cow and their convex hull

Figure 3 illustrates the flowchart for drawing the convex hull of a discrete concave curve. For a discrete concave curve containing P points, two points are iteratively selected from the

curve to draw line $_{i,j}$. If line $_{i,j}$ is tangential to the curve, then it belongs to the convex hull of this curve; otherwise, the points are discarded, and a new pair is selected and tested.

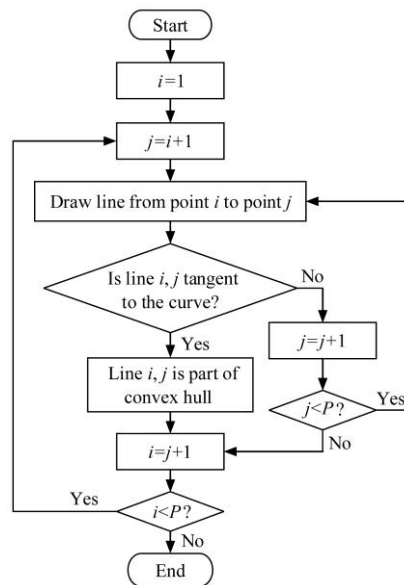


Figure 3 Flowchart of drawing the convex hull for a concave curve

The convex hull of the sacral ligaments of a cow was calculated based on the above algorithm and shown in Figure 4. The space between the convex hull and the sacral ligament indicates the potential space in which the cow can carry fat reserves. The average distance between the convex hull and sacral ligaments was calculated using Equation (1) to evaluate the visibility of the sacral ligaments quantitatively.

$$VSL = \frac{SES}{LSL} \times \frac{ASL}{WSL} \tag{1}$$

where, VSL is the visibility of the sacral ligament; SES is the area of the total space between the convex hull and the sacral ligament; LSL is the length of the sacral ligament; WSL is the width from the left hook bone to the right hook bone, and ASL is the average of WSL over the dataset; ASL/WSL is the coefficient used to eliminate the effect of individual size and shape on the VSL ; VSL is independent of the height of the cow. When the VSL is close to 0, the sacral ligament is barely visible.

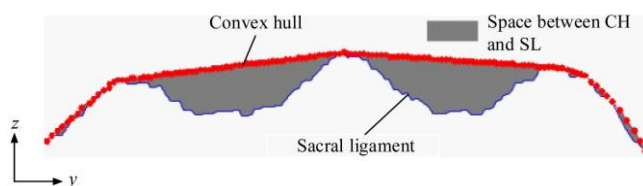


Figure 4 Space between convex hull and outline of sacral ligament

Surface curvature. The fat reserved on the hook bone and pin bone was measured by the surface curvature (SC). The bones of thin cows appear sharper than those of fat cows; therefore, the SC of the bones is larger in thin cows.

In this study, the SC of a piece of the surface was defined as the ratio of the superficial area to the shadow area of that piece of surface. Therefore, the SC is independent of the height and size of the surface. Figure 5 illustrates the hook bones of a thin cow and a fat cow. In Figure 5a, the hook bone is sharp, and the SC is 1.4; meanwhile, the hook bone in Figure 5b has a flat surface and an SC of 1.17.

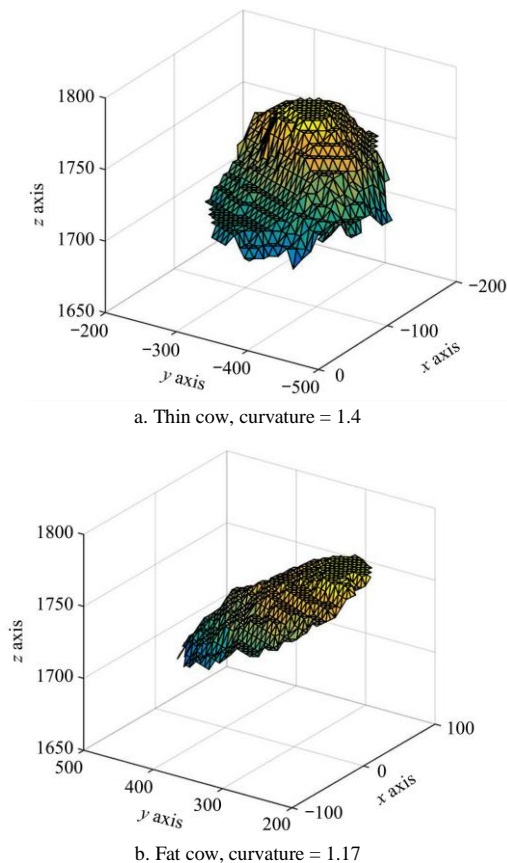


Figure 5 Surfaces of hook bones on thin cows and fat cows, with different curvature values

Back fatness. Because the 5-point BCS system mainly focuses on the area of a cow's back above the thurl, a depth threshold, denoted by DT , was used to segment the region of interest in the depth image. For the j -th column from the back to the front of the cow, the height of the point on the spine in that column was denoted by Hsp_j . The points in the depth image were filtered according to the rule described by Equation (2). The points that deviated from the Hsp_j by less than DT were reserved. Otherwise, the points were discarded. The information that was unrelated to BCS in the depth image was filtered through this process.

$$Mask_{i,j} = \begin{cases} 1 & H_{i,j} \geq Hsp_j - DT \\ 0 & H_{i,j} < Hsp_j - DT \end{cases} \quad (2)$$

where, $Mask_{i,j}$ indicates whether the point will be reserved ($Mask = 1$) or discarded ($Mask = 0$); Hsp_j is the height of the point on the spine in j -th column, and $H_{i,j}$ is the height of the point in the i th row and j -th column. DT was set as 100 in the study.

Thin cows had visible thurls with additional points classified as belonging to the thurl area. Therefore, the visibility of the thurl (VTH) can be evaluated by calculating the ratio of the area of the image after and before image cropping using Equation (3).

$$VTH = 1 - \frac{APH_{after}}{APH_{before}} \quad (3)$$

where, APH_{after} and APH_{before} are the area of the region from pin bone to hook bone after and before the image cropping, respectively. The methods for locating the pin bone and hook bone will be presented in the following sections. VTH is independent of the height and size of the cow.

2.4 Image processing

Background subtraction. The first 1200 camera depth

frames without cows were used to build the background image. The background image was then continuously updated to avoid potential error from any single background image. The number of frames for building the background was set to 1200 to ensure that the depth-frame sequence contained at least one depth frame that included the floor and walls of the scene. Background subtraction and threshold processing were performed to obtain pure depth data of cow. After background subtraction, the depth value of each pixel on the cow body was converted into the distance to the floor by adding the height of the camera, which was 3050 mm.

Image rotation. During the movement of cows, the deviation between the body axis and the horizontal axis of the image may occur, which has a great influence on the subsequent image processing. Therefore, the body axis of the cow needs to be detected and corrected. The spine of the cow was detected by calculating the highest point in each column in the depth image. A line was fitted to the group of points. The image was rotated according to the angle between the x -axis and the fitted line. The symmetry of the rotated cow was tested by calculating the overall difference between the left and right of the cow as defined by the line of symmetry^[24].

Image crop. A width threshold was used to eliminate the tail from the depth image. Measuring from the back to the front of the cow, if the width of the pixels in each column was less than a certain threshold, the pixels in this column were set as 0 to remove the tail.

Hook bone and sacral ligament detection. The contour of the back of the cow was divided into left and right parts by a symmetry line. As shown in Figure 6, points A and B , farthest away from the symmetry line, indicate the ends of the hook bones on the left and right sides of the cow's outline, respectively. The x -coordinates of points A and B indicate the x -coordinates of the hook bones. The hook bones are the bumps on the left and right sides in the curve of the sacral ligament.

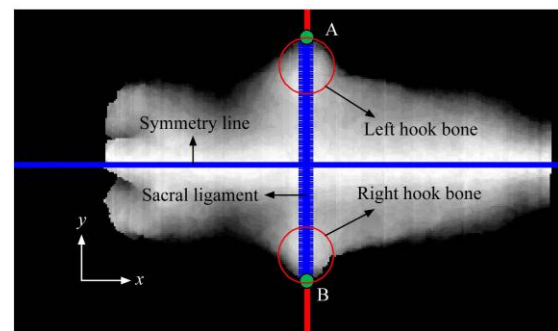


Figure 6 Detection of sacral ligaments and hook bones on cropped image

The y -coordinate and size of the bump were determined by analyzing the slope of the sacral ligament along with that of the convex hull. By combining the x -coordinate of the sacral ligament and the y -coordinate with the size of the bump, the hook bones in the depth image were detected, as shown in Figure 6. The surface curvatures of left and right hook bones were calculated and the average value of them was denoted by CHB .

The sacral ligaments were isolated by connecting points A and B and extracting the slice on the line from the depth image. After the sacral ligaments were extracted from the depth image, the visibility of it was calculated as VSL using Equation 1.

Removal of tailhead and detection of pin bones. The tailhead caused a discontinuous change in the depth image and had a major influence on the analysis of the pin bone; therefore, it was

necessary to remove the tail head from the depth image. Figure 7 illustrates the slice containing the tailhead and pin bones. As shown in Figure 7, the tailhead caused two drop points, D_1 and D_2 , in the slice. At points D_1 and D_2 , the distances between the convex hull and the slices were the maximum values, which were denoted as M_1 and M_2 , respectively. The points between D_1 and D_2 were set to 0 to remove the tailhead. From the back to the front of the cow, M_1 and M_2 decreased. If the mean of M_1 and M_2 was smaller than a predetermined threshold, the slice was defined as the end of the tailhead, as shown in Figure 7.

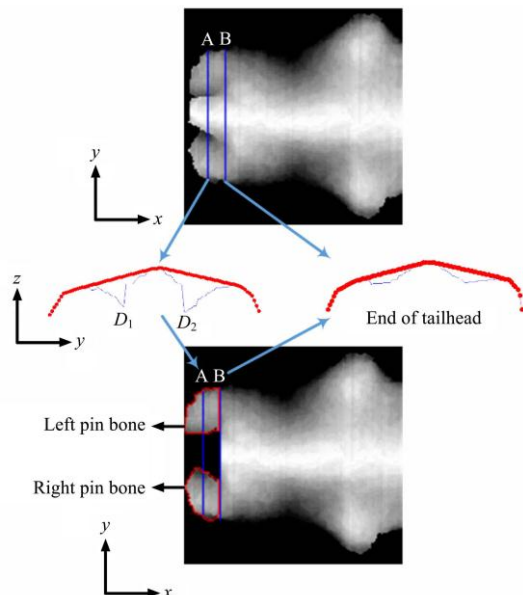


Figure 7 Detection of pin bones and removal of the tailhead by convex hull analysis

The areas separated by the removed tailhead on the left and right sides were the left pin bone and right pin bone, respectively. The surface curvatures of left and right pin bones were calculated and the average value of them was denoted by CPB . For fat cows, the pin bones are entirely hidden, and the length of the tailhead is dramatically shorter than in thin cows. In this study, if the length of the tailhead of a cow was less than 50 mm, then the pin bones of this cow were considered entirely hidden, and CPB was set as 1. For a specific cow, before any further calculations, the length of the tail head was multiplied by ASL/WSL , which was defined in Equation (1), to eliminate the effect of individual cow size.

2.5 Prediction models

For each selected depth image, the image processing generated four features (i.e. VTH , VSL , CHB , and CPB) according to the methods described above. The distribution of those four features among BCS classes were analyzed by plotting their boxplot, and analysis of variance (one-way ANOVA) was used to test for differences in measured features with regard to the BCS class. The decision tree learning method was used to build a classification model to predict BCS values based on features. Linear regression and a backpropagation (BP) neural network were used to build regression models for continuous BCS evaluation. The features and the BCS values were normalized to a range of 0 to 1 before they were applied to the regression models.

Decision tree learning. Decision tree learning^[26] was utilized to classify the four feature variables into seven BCS levels (2.25 to 3.75 with 0.25 interval) according to the given scores. This predictive model maps a group of observations of an item to the target value of the item. In these tree structures, leaves

represent class labels and branches represent conjunctions of features that lead to those class labels by using if-then rules. The goal of training is to find the optimal criteria for the if-then rule in each branch of the tree. In the prediction phase, the trained model takes several variables from one observation as input, and a path from the root node to a leaf node is determined by comparing the variables and the criteria of the if-then rule in each branch. The class label that the leaf node represents is the output of the model.

Linear regression. The linear regression model operated by assuming that the larger the four features were, the thinner the cow was. When the four features were all 0, the cow was expected to have had a high body condition score. It was assumed that the human evaluator scored the cow based on the highest body condition score in the herd and then reduced the score by perceiving the sharpness of the thurl, sacral ligament, hook bone, and pin bone. Based on that assumption, a model for BCS regression was designed as follows:

$$BCS = \mu - w_1 \times VTH - w_2 \times VSL - w_3 \times CHB - w_4 \times CPB \quad (4)$$

where, μ was the highest score in the herd and w_1 , w_2 , w_3 , and w_4 were the scores that were subtracted from μ due to the sharpness of the thurl, sacral ligament, hook bone, and pin bone, respectively. VTH and VSL are the visibility of the thurl and sacral ligament, respectively. CHB and CPB are the curvature of the hook bone and pin bone, respectively.

BP neural networks. The BP neural network is a multilayer feedforward network that is trained according to an error BP algorithm. This method is one of the most widely used neural network models^[27]. A BP network can be used to learn and store mapping relations in an input-output model, and there is no need to disclose the mathematical equation that describes these mapping relations^[28]. Due to this characteristic, the BP network is a feasible way to regress the relationship between the inputs and output, regardless of whether it is linear or nonlinear. In the network in this study, the input layer contains four neurons for each feature and two hidden layers with five neurons in each one. The output of the model is the body condition score that is predicted by the input features. The neurons were fully connected to each other in the different layers. The transfer functions of the hidden layers and output layer were 'tansig' and 'purelin' respectively. The maximum epoch of training was set as 100. The learning rate and training goal were 0.1 and 0.0004, respectively.

2.6 Model Evaluation

The decision tree model was evaluated on the entire dataset (D_E) and the training dataset (D_T) with 3-, 5-, and 10-fold validation. K -fold validation was not applicable to the D_E because the numbers of samples in different classes of the D_E were not even. The linear regression and BP network model were built based on the D_T and tested on the D_E . The three models were evaluated by the mean absolute error (MAE), the rate of correct classifications (only for the decision tree model), and the rates of predicted scores within 0.25 and 0.5 BCS points of manual scores. The correlation (R^2) between the results of the model and the target BCS values was also calculated to evaluate the two regression models.

3 Results

3.1 Correlations and distribution of features

Correlations were calculated between BCS and the four features (visibility of thurl, visibility of sacral ligament, curvature of hook bone, and curvature of pin bone) and are shown in Table 2. All four features were negatively correlated with BCS ($p < 0.01$). The visibility of the sacral ligament ($r = -0.86$) had the strongest

correlation with BCS, followed by the curvature of the pin bone ($r=-0.85$) and the curvature of the hook bone ($r=-0.75$). The visibility of the thurl ($r=-0.73$) had the weakest correlation with BCS among the four predictive features. The four features were positively correlated with each other ($p<0.01$). Among the features, the correlation between the visibility of the sacral ligament and the curvature of the pin bone was the strongest ($r=0.74$), while the correlation between the visibility of the thurl and the curvature of the hook bone was the weakest ($r=0.50$). Other correlations among the features ranged from 0.62 to 0.70.

Table 2 Correlations among BCS and the features of the thurl, sacral ligament, hook bone, and pin bone

Item	BCS	VTH	VSL	CHB	CPB
BCS	1.00	-0.73	-0.86	-0.75	-0.85
VTH	-0.73	1.00	0.70	0.50	0.67
VSL	-0.86	0.70	1.00	0.64	0.74
CHB	-0.75	0.50	0.64	1.00	0.62
CPB	-0.85	0.67	0.74	0.62	1.00

Note: All p -values were <0.01 .

Figure 8 illustrates the distribution of the features among BCS levels in the entire dataset (D_E). As shown in the figure, all the features gradually decreased as the BCS values increased of cows.

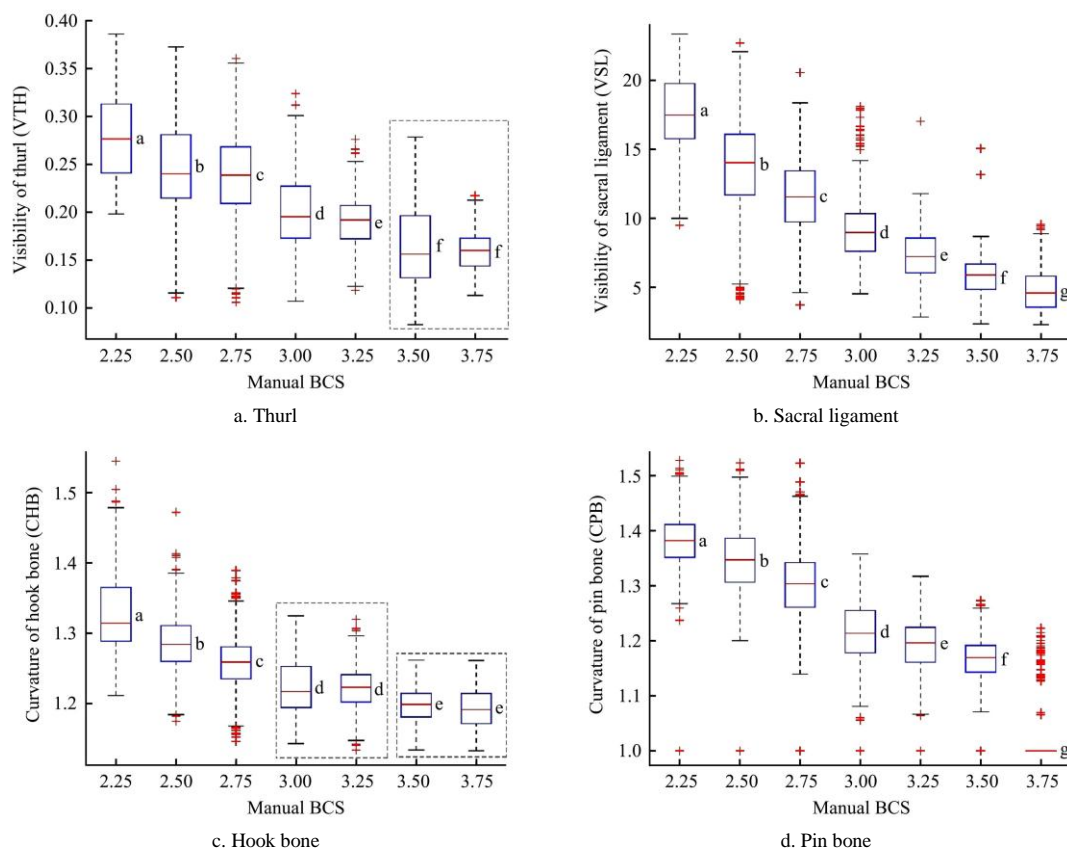


Figure 8 Distribution of features by BCS. Different letters indicate data groups being significantly different ($P < 0.001$)

Table 3 Classification results of the decision tree learning model based on different validation data

Validation data	Correct classification/%	≤ 0.25 BCS points/%	≤ 0.5 BCS points/%	MAE
3-fold (D_T)	57.67	94.85	99.47	0.12
5-fold (D_T)	58.72	94.22	99.47	0.12
10-fold (D_T)	62.92	95.38	99.68	0.11
Entire data (D_E)	61.81	95.48	99.65	0.11

Most features among BCS levels are significantly different between grouped data except for VTH values in BCS 3.5 and 3.75, CHB values in BCS 3 and 3.25, as well as CHB values in BCS 3.5 and 3.75. Two groups of data cannot be distinguished from each other if they are not significantly different. However, other features in these BCS groups (i.e. VSL and CPB) are significantly different, and provide great variability that can classify and predict these BCS values. Distributions and medians of VTH values in BCS 2.5 and 2.75 are close. However, the result of ANOVA test shows they are significantly different and come from the different normal populations, which means their variance is quite different and can provide variability for classification and regression. Compared to the other features, the VSL showed an improved linear relationship with the BCS and reduced interclass overlap. The VSL of a fat cow ranges from 2 mm to 6 mm, which indicates that the sacral ligament is barely visible.

The CHB and pin bones ranged from 1 to 1.5. The CHB dropped sharply as BCS increased from 2.25 to 2.75, while the tendency became flat for BCS values greater than 3, which showed that there was a nonlinear relationship between the CHB and BCS. The CPB had a similar tendency for BCS values from 2.25 to 2.75 and those from 3 to 3.5. However, the drop from 2.75 points to 3 points was considerable.

3.2 Results of decision tree learning

Table 3 illustrates the classification results of decision tree learning using 3-, 5-, and 10-fold cross-validation based on the training (D_T) and the entire dataset (D_E). As shown in Table 3, the accuracies of 3-, 5-, and 10-fold cross-validations were similar, but the model achieved the highest accuracy when using 10-fold cross-validation, with which 95.38% and 99.68% samples were classified within 0.25 and 0.5 BCS points, respectively, of the manual scores. The accuracy of classifications was improved by

5.25% by increasing the K value from 3 to 10. When the decision tree was trained on D_T and tested on D_E , the result was not significantly different from that with 10-fold cross-validation on D_T .

3.3 Results of regression

The BCS regression model was fitted to a training dataset (D_T), and the result was as follows:

$$BCS = 3.94 - 0.35 \times VTH - 0.71 \times VSL - 0.67 \times CHB - 0.72 \times CPB \quad (5)$$

The parameters in the model showed that the highest theoretical score in the herd was 3.94 and that a sharp thurl, sacral ligament, hook bone, and pin bone will reduce the BCS by 0.35, 0.71, 0.67, and 0.72, respectively. The R^2 of the regression was 0.88 ($p < 0.01$) as shown in Table 4. The BP network took 12 epochs to finish the training on the D_T . The linear regression and BP network were tested on D_T and the D_E ; the results are shown in Table 4.

For the same model, the rates of predicted scores within 0.25 and 0.5 BCS points were not significantly different when using D_T and D_E as testing data. However, the R^2 was reduced by 0.08 and 0.07 in the linear regression model and BP network, respectively, when they were tested on D_E . In general, the BP network achieved higher accuracy than linear regression regarding all the

performance indicators, especially the proportion of results within 0.25 BCS points of manual scores; in this respect, the BP network was over 5% more accurate than linear regression.

Table 4 Performance of the linear regression model and BP network

Model & testing dataset	≤ 0.25 BCS points/%	≤ 0.5 BCS points/%	MAE	R^2
Linear regression (D_T)	85.61	99.58	0.14	0.88
Linear regression (D_E)	86.14	99.69	0.14	0.80
BP network (D_T)	92.65	99.68	0.12	0.91
BP network (D_E)	91.68	99.88	0.12	0.84

The correlations between the camera and manual BCS values were analyzed to evaluate the regression models; the results are shown in Figure 9. The two regression models were similar in overall performance, but the predicted scores from the BP network were more concentrated than those from linear regression. Positive drift of the predicted scores versus the manual scores was found at BCS 3 in the linear regression model with D_T and D_E , as well as BCS 2.25 in the BP network model. The predicted scores for BCS 3.75 showed negative drift in both models with D_T and D_E .

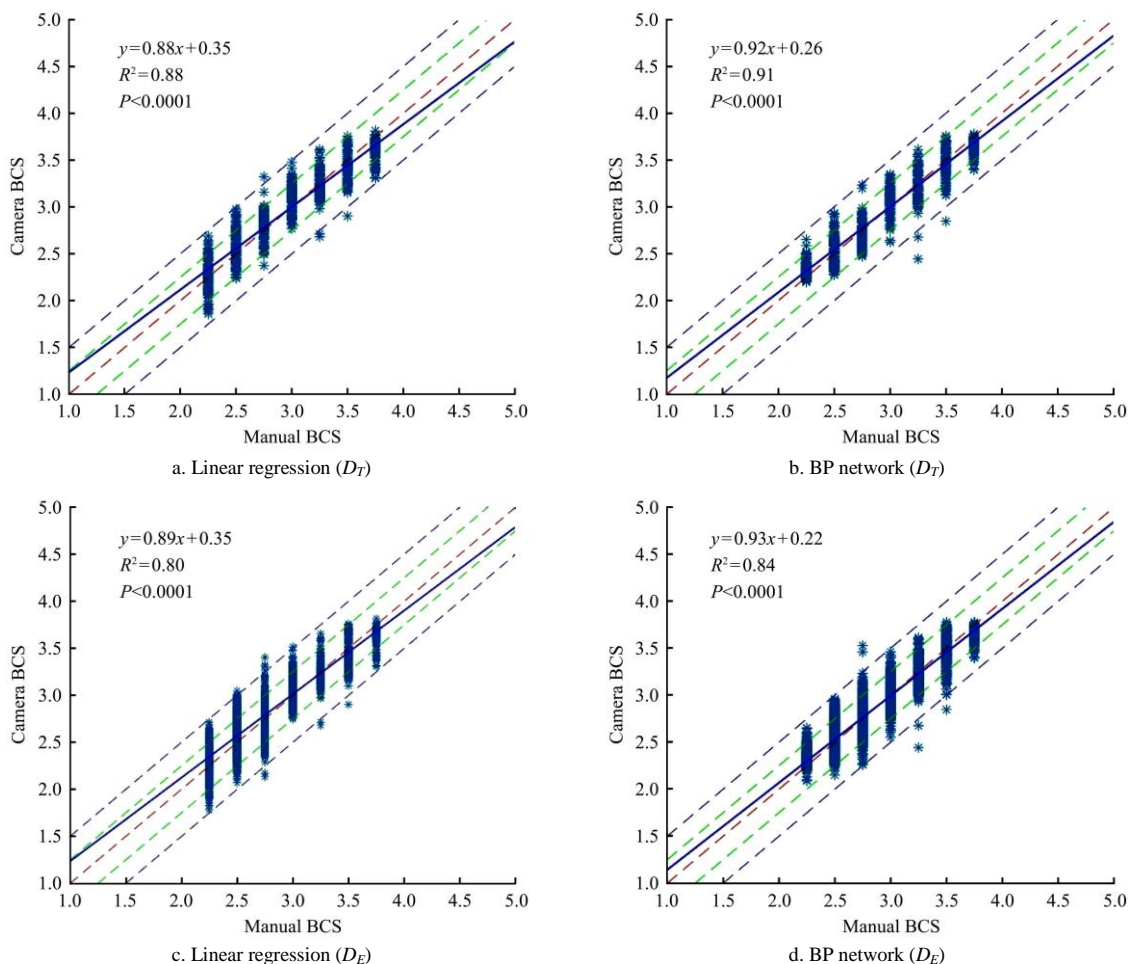


Figure 9 Predicted BCS versus human BCS

3.4 Predicted scores for individual cows

Figure 10 illustrates the MAE and standard deviation of the BCS error of each cow during the study period (sorted by MAE) when the BP network was used as a prediction model. The average MAE was 0.11, and all cows had MAEs lower than 0.25. The average standard deviation was 0.069, where 95% of the SD values of cows were less than 0.1.

Figure 11 illustrates four patterns of camera scores versus manual scores in four selected cows. In Figure 11a, the predicted scores were close to the manual score during the study period, and the MAE was 0.02. The result in Figure 11b was the most common case among cows, where all the intervals were less than 0.25 and the average MAE of the cow was 0.13. Figure 11c illustrates the cow with the maximum MAE (0.22) in Figure 10.

In Figure 11d, the cow had an abortion, and its BCS dropped from 3 to 2.5 in the first 21 days of lactation 2; the predicted scores

tracked that change successfully and even found the change earlier than the manual score changed, with an MAE of 0.09.

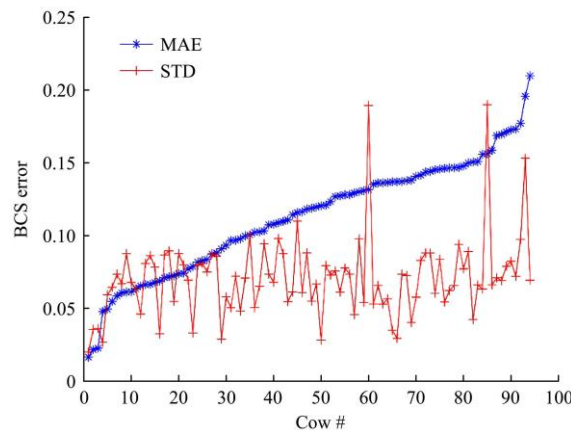


Figure 10 MAE and standard deviation of the BCS predicted by the BP network for each individual cow (sorted by MAE)

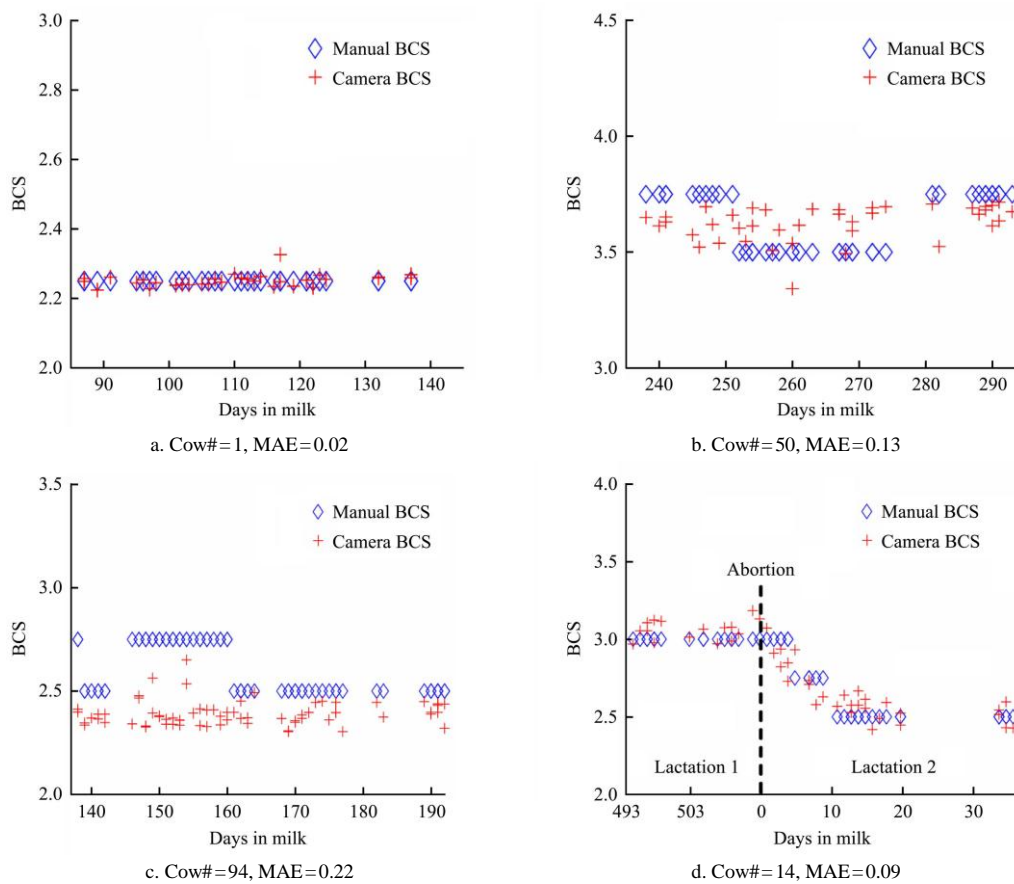


Figure 11 Different patterns of camera scores versus manual scores among cows

3.5 Discussion

Our work describes a method for scoring the body condition of cows with high accuracy. The system was developed and tested based on longitudinal data (2 months) from 94 cows. Four features were extracted from depth images, and three models were used to predict the scores based on these features. Compared to previous studies, this study improved the percentage of predicted scores within 0.25 BCS points of manual scores, raising this proportion to 90% using a decision tree and a BP network model with a fully automatic system.

The four features identified in this study all have strong positive correlations with BCS. A linear model was built to demonstrate the weights of the features when the human assigned the scores. The model partly explained the human scoring

procedure and demonstrated the theoretical highest score in the herd. Due to the nonlinear characteristics of the features, the accuracy of the linear model was lower than that of the BP network model, which is able to regress linear and nonlinear relationships. The results showed that 95.48% of the samples were scored within 0.25 BCS points of the manual score using a decision tree learning model, which is a higher accuracy rate than that of the linear regression model (86.14%) or the BP network (91.68%). This demonstrates that the decision tree is more feasible for classification than other models if the data are treated as categorical.

In this study, the hook bone and pin bone were detected, and their curvatures were used as indicators of fatness. Bewley et al. (2008) used 2D digital images to analyze the outline of the backs of

cows and generate hook and tailhead descriptions related to BCS. The results showed that the hook angle, posterior hook angle, and tailhead depression were significant predictors of BCS. In our study, angle descriptions were not involved in the model because the edge of the entity in the depth image had a great influence on the 3D points that were close to the outline, which reduced the accuracy of the angles calculated from the outline in the 3D image. The descriptions of the sacral ligament and thurl were explored in this study because these two areas are frequently evaluated in the existing BCS system^[14].

Additional researchers also developed automatic BCS systems based on 2D^[16,19] and thermal^[17,18] image processing technology. However, it is difficult to accurately detect specific anatomical points of a cow and extract fatness-associated features closely related to BCS based on two-dimensional image data alone. The 3D images used in this study not only provided depth information but also made it possible to measure the physical traits of the cow without additional image calibration. Weber, et al.^[20] developed an automatic 3D optical system with which to estimate the backfat thickness (BFT) of cows. The correlation between the observed and estimated BFT values was 0.96, which demonstrated the feasibility of using 3D images to measure the body fat reserves of a cow. It would be worthwhile to study the relationship between BFT and BCS to build a model predicting the latter from the former. Fischer et al.^[21] used a 3D camera to capture the surface of the hindquarters of cows, and four anatomical landmarks were identified manually from the surface to predict BCS. However, that study still involved manual processing. Spoliarsky et al. also used 3D images of backs to calculate the relative heights of different parts of the cows, and the height information was combined with weight and age to build a model with which to predict BCS. In our study, the model was based only on depth images and required no additional information, which makes this system easy to implement in commercial applications.

Sandgren and Emanuelson^[23] reported the validation results from a commercial camera system; a total of 95% of the cows were scored within 0.25 BCS points of manual scores, with 99% of the scores having a standard deviation of less than 0.1. The result of our study showed that all cows were scored within 0.25 BCS points compared to the manual scores, and 95% of the cows had a standard deviation of less than 0.1. Therefore, our system exhibited similar performance to the reported commercial camera. However, the commercial camera used daily rolling average scores from seven-day periods as the output, which can improve the consistency of the output scores. Rolling average operation will greatly reduce the dynamic tracking performance of the scoring system, which makes the system insensitive to abnormal changes in a short time. The system proposed in this paper can ensure high precision and good tracking performance without rolling average.

The manual scores were categorical data. However, the fatness of a cow may fall between two categorical values. Therefore, the difference between the predicted scores and manual scores may be caused by the difference between the actual BCS and the manual BCS in some cows. A prior study also showed that the MAE of a well-trained expert was 0.25. Thus, the standard deviation of the predicted scores could be a good indicator for evaluating the performance of the automatic BCS system when the MAE of the system is lower than 0.25.

The region of the short ribs is another anatomical feature associated with the fatness of a cow. This area was ignored in the current study because the end of the short rib area is invisible on a

fat cow, which made it difficult to determine and analyze that area. Future studies should focus on detecting the short rib area and analyzing the fat reserved on it to further improve the accuracy of the system.

4 Conclusions

Specific areas that are related to BCS, including the thurls, sacral ligaments, hook bones, and pin bones, can be accurately located through depth-image processing and the use of convex hulls. Measured values of the visibility and curvature of the four areas were strongly correlated with manually assigned BCS values. When the BP network was used, the system can score each cow within 0.25 points compared to the target BCS during the two-month study period (i.e. MAE of each individual cow is less than 0.25); meanwhile, the averaged MAE and SD of all cows were 0.11 and 0.069, respectively. The result shows that the system has high precision and good tracking performance, which demonstrates that the automatic system has the possibility of being more accurate than human scoring. Future studies should focus on analyzing the short rib area to provide an additional fatness-related feature and further improve accuracy.

Acknowledgements

The authors acknowledge the Coldstream farm, University of Kentucky, the US for their cooperation in the project. The authors also acknowledge Amanda Stone for scoring the body condition of cows during the data collection. The work was sponsored by the Key R&D and Promotion Projects in Henan Province (Science and Technology Development, No. 192102110089), Open Funding Project of Key Laboratory of Equipment and Informatization in Environment Controlled Agriculture, Ministry of Agriculture and Rural Affairs, P.R. China (No. 2011NYZD1804), and Key Scientific Research Project Plan of Colleges and Universities in Henan Province (No. 19A416003).

[References]

- [1] Grainger C, Wilhelms G D, McGowan A A. Effect of body condition at calving and level of feeding in early lactation on milk production of dairy cows. *Australian Journal of Experimental Agriculture*, 1982; 22(115): 9–17.
- [2] Enevoldsen C, Kristensen T. Estimation of body weight from body size measurements and body condition scores in dairy cows. *Journal of Dairy Science*, 1997; 80(9): 1988–1995.
- [3] Stockdale C R. Body condition at calving and the performance of dairy cows in early lactation under Australian conditions: a review. *Australian Journal of Experimental Agriculture*, 2001; 41(6): 823–839.
- [4] Berry D. P, Macdonald K A, Penno J W, ROCHE J R. Association between body condition score and live weight in pasture-based Holstein-Friesian dairy cows. *Journal of Dairy Research*, 2006; 73(4): 487–491.
- [5] Russel A J F, Doney J M, Gunn R G. Subjective assessment of body fat in live sheep. *The Journal of Agricultural Science*, 1969; 72(3): 451–454.
- [6] Roche J R, Friggens N C, Kay J K, Fisher M W, Stafford K. J, Berry D P. Invited review: Body condition score and its association with dairy cow productivity, health, and welfare. *J Dairy Sci*, 2009; 92(12): 5769–5801.
- [7] Bewley J M, Schutz M M. Review: An interdisciplinary review of body condition scoring for dairy cattle. *Prof. Anim. Sci.*, 2009; 24: 507–529.
- [8] Waltner S S, Mcnamara J P, Hillers J K. Relationships of body condition score to production variables in high producing Holstein dairy cattle. *J Dairy Sci*, 1993; 76(11): 3410–3419.
- [9] Heuer C, Schukken Y H, Dobbelaar P. Postpartum body condition score and results from the first test day milk as predictors of disease, fertility, yield, and culling in commercial dairy herds. *Journal of Dairy Science*, 1999; 82(2): 295–304.
- [10] Pryce J E, Coffey M P, Simm G. The relationship between body

- condition score and reproductive performance. *Journal of Dairy Science*, 2001; 84(6): 1508–1515.
- [11] Randall L V, Green M J, Chagunda M G G, Mason C, Archer S C, Green L E, et al. Low body condition predisposes cattle to lameness: An 8-year study of one dairy herd. *Journal of Dairy Science*, 2015; 98(6): 3766–3777.
- [12] Bewley J M, Schutz M M. An interdisciplinary review of body condition scoring for dairy cattle. *The professional animal scientist*, 2008; 24(6): 507–529.
- [13] Bewley J M, Boehlje M D, Gray A W, Hogeveen H, Kenyou S J, Eicher S D, et al. Assessing the potential value for an automated dairy cattle body condition scoring system through stochastic simulation. *Agricultural Finance Review*, 2010; 70(1): 126–150.
- [14] Elanco. The 5-point body condition scoring system. Available: <https://www.elanco.us/pdfs/ai10752-body-condition-score-insert.pdf>.
- [15] Bewley J M, Peacock A M, Lewis O, Boyce R E, Roberts D J, Coffey M P, et al. Potential for estimation of body condition scores in dairy cattle from digital images. *J Dairy Sci*, 2008; 91(9): 3439–3453.
- [16] Azzaro G, Caccamo M, Ferguson J D, Battiato S, Farinella G M, Guamera G C, et al. Objective estimation of body condition score by modeling cow body shape from digital images. *J Dairy Sci*, 2011; 94(4): 2126–2137.
- [17] Halachmi I, Polak P, Roberts D J, Klopčič M. Cow body shape and automation of condition scoring. *J Dairy Sci*, 2008; 91(11): 4444–4451.
- [18] Halachmi I, Klopčič M, Polak P, Roberts D J, Bewley J M. Automatic assessment of dairy cattle body condition score using thermal imaging. *Computers and Electronics in Agriculture*, 2013; 99: 35–40.
- [19] Bercovich A, Edan Y, Alchanatis V, Moallem U, Parmet Y, Honig H, et al. Development of an automatic cow body condition scoring using body shape signature and Fourier descriptors. *J Dairy Sci*, 2013; 96(12): 8047–8059.
- [20] Weber A, Salau J, Haas J H, Junge W, Bauer U, Harms J, et al. Estimation of backfat thickness using extracted traits from an automatic 3D optical system in lactating Holstein-Friesian cows. *Livestock Science*, 2014; 165: 129–137.
- [21] Fischer A, Luginbuhl T, Delattre L, Delouard J M, Faverdin P. Rear shape in 3 dimensions summarized by principal component analysis is a good predictor of body condition score in Holstein dairy cows. *J Dairy Sci*, 2015; 98(7): 4465–4476.
- [22] Delaval. DeLaval body condition scoring BCS. Available: <http://www.delavalcorporate.com/our-products-and-services/farm-support/delaval-body-condition-scoring-bcs/>.
- [23] Sandgren C H, Emanuelson U. Consistency of measurements from an automatic body condition scoring camera. *Precision Dairy Farming 2016*, Leeuwarden: Wageningen Academic Publishers, 2016; pp. 285–290.
- [24] Spoliansky R, Edan Y, Parmet Y, Halachmi I. Development of automatic body condition scoring using a low-cost 3-dimensional Kinect camera. *J Dairy Sci*, 2016; 99(9): 7714–7725.
- [25] Alvarez J R, Arroqui M, Mangudo P, Toloza J, Jatip D, Rodríguez J M, et al. Body condition estimation on cows from depth images using Convolutional Neural Networks. *Computers Electronics in Agriculture*, 2018; 155: 12–22.
- [26] Utgoff P E, Berkman N C, Clouse J A. Decision tree induction based on efficient tree restructuring. *Machine Learning*, 1997; 29: 5–44.
- [27] Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998; 86(11): 2278–2324.
- [28] Yao X. Evolving artificial neural networks. *Proceedings of the IEEE*, 1999; 87(9): 1423–1447.