Original papers

Segmentation of body parts of cows in RGB-depth images based on template matching

Nan Jia\textsuperscript{a}, Gert Kootstra\textsuperscript{b}, Peter Groot Koerkamp\textsuperscript{b}, Zhengxiang Shi, Ph.D., Professor\textsuperscript{a}, Songhuai Du\textsuperscript{a,*}

\textsuperscript{a} China Agricultural University, Beijing, China
\textsuperscript{b} Wageningen University, Farm Technology Group, Wageningen, The Netherlands

A R T I C L E  I N F O

Keywords:
Body parts segmentation
RGB-depth images
Computer vision

A B S T R A C T

Body cleanliness is considered an important indicator for evaluating cow welfare. At present, assessing the cleanliness of different cow body parts is considered as a subjective and labor-intensive task. Automatic body cleanliness scoring needs to start with body parts segmentation. Despite the fact that pattern recognition methods for human body detection and analysis have flourished in the last decade, computer vision analysis of cow body-part is scarce in literature, and most of cow body detection in video images recognizes and segments the cow as a whole, not using all of the body part information. This study presents a computer vision method that automatically identifies nine cow’s body parts, i.e. head, torso, udder, belly (or rear), left foreleg, right foreleg, left hindleg right hindleg and tail. We used a method for body parts detection using an improved version of template matching. The entire image process included video recording, image restoration and pre-processing, skeleton extraction, templates matching, recognition of right or left leg, calibration and body parts identification. 1421 side-view images of 113 cows and 859 back-view images of 75 cows were collected on a Chinese research dairy farm using an RGB-depth camera installed in the barn for Holstein lactating cows. The average body parts segmentation accuracies for side-view and back-view were 96% and 91%, respectively. The results indicate that it is possible to automatically detect and extract body parts from RGB-depth images without any manual interference.

1. Introduction

An important indicator for the welfare of cows is the cleanliness of the cow, as it is one of the critical factors influencing bacterial contamination of milk (Zucali et al., 2011), somatic cell count (SCC) and subclinical intra-mammary infection rate (Schreiner and Ruegg, 2003). The cleanliness also has economic consequences, as it influences milk quality, milk quantity and reproduction (Ellis et al., 2007). Cow cleanliness scoring is typically done using a 4 or 5 points scoring method ranging from clean to very dirty (Hultgren and Bergsten, 2001; Reneau et al., 2005; Lombard et al., 2010; Eckelkamp et al., 2016). For the assessment, the cleanliness of body parts (e.g. leg, thigh, belly, flank, rear and udder) is normally assessed separately.

In current practice, cow cleanliness is assessed manually by experts. As this is a time-consuming and expensive procedure, limited number of assessments are performed. Human scoring, moreover, is subjective and inconsistent. To improve both the quantity as well as the quality of cleanliness scoring, there is a strong need for automated scoring systems using camera systems. The development of automatic cleanliness evaluation needs 2 steps: (1) segmentation of the different body parts in the photographs and (2) detection of dirt on these segments and calculation of the body hygiene score. In this paper, we focus on the first step, the segmentation of body parts of the cows in color-depth (RGB-D) images, as this is a challenging and currently unsolved problem.

Computer-vision methods to segment different cow body-parts for cleanliness scoring are scarce in the literature. Here, we discuss related work concerning the use of computer vision to automate monitoring of cattle, as well as some relevant studies on body-part detection on humans and other animals.

Pezzuolo et al. (2018) evaluated the use of multiple low-cost structured-light depth cameras for several 3D body measurements of dairy cows and reported uncertainty ranging between 3 and 15 mm. For the
autonomous monitoring of body weight, Song et al. (2018) developed a method to measure several morphological traits from top-view depth images of the cow. Using these traits in combination with additional information on milk production and parity, the body weight was estimated. Other studies estimate the body condition score (BCS), a commonly used indicator of the nutritional status in beef and dairy cows. In a review, Alvarez et al. (2017) concluded that automatic systems based on image processing can reach human-level performance. Bercovich et al. (2013), for instance, developed an automatic BCS using body shape signature and Fourier descriptors.

Image processing also allows for behavior analysis. Leonard et al. (2019) adopt an overhead Kinect monitored and classified cow postures in farrowing stalls. Cangar et al. (2008), for instance, developed an automatic image-analysis system to identify the locomotion and posture behavior of pregnant cows prior to calving. Their method records position and body-size measurements over time in order to classify behavior as, e.g., standing or lying, and eating or drinking. Porto et al. (2013) developed a multi-camera system to detect cow lying behavior by training the Viola-Jones object-detection algorithm (Viola and Jones 2001). Ahn et al. (2017) proposed a method for the detection of cow mounting-behavior to determine the optimal time of insemination by using the support vector machine (SVM) classifier with motion history image feature information. To analyze cow locomotion, Song et al. (2018) developed a vision-based system to detect lameness by tracking the locations of the hooves over time. They reported strong correlations with human lameness scored using the timing of the placements of the hooves. Pluk et al. (2012) studied on the automatic measurement of touch and release angles of the fetlock joint for lameness detection in dairy cattle using a combination of computer vision and a pressure mat, further improved in (Romanini et al., 2013). Viazzi et al. (2013) developed the body-movement-pattern score, which uses the back posture to classify lameness.

Most of the existing work on body-part detection focuses on human posture recognition from 2D or 3D images. Felzenszwalb and Huttenlocher (2005) presented a computationally efficient framework for part-based modeling and recognition of objects using pictorial structure models, and demonstrated the technique on the detection of facial features and human body parts. Bourdev and Malik (2009) studied the detection, segmentation and pose estimation of people in images based on the detection of so-called poselets using histograms-of-orientation-gradients (HOG) images features in combination with a Support Vector Machine (SVM) classifier. Simo-Serra et al. (2012) estimated 3D body pose unambiguously by imposing kinematic constraints. Shotton et al. (2011, 2013) used a Random Forests (RF) classifier with the global 3D centers of probability mass for each body part using simple depth image to locate the positions of 3D skeletal joints. Currently, convolutional neural networks become popular for the detection of body keypoints for the detection of human pose, e.g. in (He et al., 2017). Li et al. (2019) applied dominant orientation templates and brightness ratio templates to detect group-housed pigs.

Computer-vision methods for the detection of body parts of animals, on the other hand, are scarce in literature. Piscocchi et al. (2014) applied Structural Support Vector Machine (SSVM) to identify the body-parts of animals using both 2D and 3D features, tailored for dogs independently from the size and breed. Zhao et al. (2017) applied RF decision to segment eight body regions, i.e., head, neck, body, forelegs, hind legs and tail with local binary pattern (LBP)-based depth features to analyze cow posture. This method concentrates on the detection of the different body parts from other viewpoints, necessary for a proper evaluation of the cow cleanliness. To assess cow behavior, Jiang et al. (2019) applied FLYOLOv3 deep learning to detect key parts (e.g. head, trunk and legs) of dairy cow. The accuracy was 99.18% and the average detection rate was 94%.

A promising method to segment the different cow body parts is to describe the cow’s body using a skeleton (or medial axis). Skeletonizing is an effective method to represent and analyze object shapes (Saeed et al., 2010). Shape similarity based on skeleton matching usually performs better than using the full contour or other shape descriptors in the presence of partial occlusion, noise, and articulation of parts (Borgefors et al., 2001). Based on the image skeleton, a skeleton graph can be constructed. Bai et al. (2008) used such skeleton graphs and proposed a robust object detection and segmentation method based on Skeleton Path Similarity Matching (SPSM). Liu et al. (2014) adopted this method to describe and segment images of cows into a number of body parts for the analysis of their gait.

In this paper, we propose a method for body-part segmentation based on template matching for endpoint classification using color-depth (RGB-D) images. We apply the method for the detection of nine body-parts of dairy cows in an indoor environment. The novelty of our approach is twofold. Firstly, unlike the existing works, which only recognizes the whole body or seven body parts without udder from side views, our approach can detect nine different body parts, e.g. head, torso, belly, udder, left front leg, right front leg, left hind leg, right hind leg and tail from different viewpoints of the cow(side and back). We propose a color-depth based method for cow body segmentation identifying cows’ body parts. Secondly, we applied a template-based method for endpoints classification, Skeleton Path Similarity Matching (SPSM) (Bai et al.2008), which provided a robust description of the posture of cows, despite the variation in their appearance and relative pose towards the camera.

2. Materials and methods

2.1. Data acquisition

All data were collected for 3 consecutive days at the Sifang Experimental Dairy Farm in Datong, China, July 1 to July 3, 2017. From the 2000 cows on the farm, 113 lactated Holstein cows were collected for body-part detection. Color-and-depth images have been taken using the Microsoft Kinect™ V2 RGB-D camera. The Kinect V2 provides a depth frame with the resolution of 512 x 424 pixels and a color frame with 1920 × 1080 pixels. The camera was connected to a personal computer and operated using the Kinect for Windows Software Development Kit (SDK) 2.0.

A total number of 5070 RGB-D images were taken, of which 2680 (52.86%) images containing a valid standing or walking cow were manually selected. The other images were rejected for different reasons: (1) images not containing a cow’s side or back (33.33%, 1690 images), (2) images with incomplete depth data of the cow body due to direct sunlight from the roof gap (7.51%, 381 images), (3) images that contained severe occlusions from another cow (3.00%, 152 images), and (4) images with rapid movement resulting in blurred images (3.3%, 167 images). We performed the selection of images manually, as the focus of this paper is on the segmentation of cow body parts. In a future system, however, the image acquisition should be fully automated. We envision to put the camera at the exit of milking parlor, providing a standardized pose of the animal, preventing occlusions by other cows, as the animals pass there one at a time, and providing a completely controlled environment without sun light interference. With the increasing light sensitivity of future cameras, exposure time can be reduced, preventing image blur due to rapid movements.

The accepted data can be separated into four main cases: (a) complete data side view (645 image), (b) complete data back view (303 image), (c) some missing data due to some body movement (97 image), (d) some missing data due to sunlight inference (376 image). Examples of accepted and rejected images in the different categories can be observed in Fig. 1.

2.2. Image processing

The purpose of the developed image-processing method is to recognize and segment rear- and side-view images of cows into eight different
body-part, i.e., (i) head, (ii) torso, (iii) left-front leg, (iv) right-front leg, (v) left-rear leg, (vi) right-rear leg, (vii) tail, and (viii) udder. Fig. 2 gives an overview of the proposed method. The workflow consists of a number of steps: (1) restoration, (2) image segmentation, (3) image selection, (4) skeletonization, (5) skeleton classification and (6) body-part segmentation. The different steps are described in the following subsections.

2.2.1. Depth-image restoration

The depth image obtained by RGB-D camera contains random noise. For instance, the cow’s head and frunk which are shown in Fig. 3 (a) red box are accompanied by a lack of depth information. Therefore, the information lost in the depth image (the hole on the cow depth map) needs to be repaired and filled. The image restoration is for the entire depth frames. The depth images contain some missing data. To restore the depth data, two steps are performed. In the first step, the depth images are spatially filtered using a majority (or mode) filter:

$$D_t'(p) = \text{mode}\{ D_t(q) | q \in N_{5 \times 5}(p) \}$$  (4)

where $D_t(p)$ is the original depth value of pixel $p$ at time $t$, $N_{5 \times 5}(p)$ is the neighborhood function that gives the set of all pixels in the $5 \times 5$ neighborhood of $p$, and $\text{mode}(\cdot)$ returns the mode of the set $S$, i.e., the majority count of the elements in $S$. In the second step, the depth images are filtered in the temporal dimension using a weighted averaging filter:

$$D_t''(p) = \frac{1}{N} \sum_{t'=0}^{N} D_{t'}(p)$$  (5)

where we used $N = 5$ in our experiments. Fig. 3 gives an illustration of the depth-image restoration. The image restoration method makes the boundary of the animal with respect to the background smoother, by applying the mode-filter and temporal averaging filter. As we use rather small kernel sizes, the shape of the animal is not adapted significantly.

2.2.2. Cow segmentation

To segment the cow from the background, we make use of a background depth-image, as depicted in Fig. 4 a. This image contains the same scene, but only with stationary objects, that is without cows or human beings. Directly subtracting the background depth-image, $D_b$, from the current restored depth image, $D_t''$, results in poor cow segmentation due to noise and small displacements of the camera. We therefore developed a more robust background-subtraction method based on the histograms of the aforementioned depth images, $H_b(k)$ and $H_r(k)$ (see Fig. 4 c and 4d). The method consists of two steps; (1) determining the depth range in which the cow is positions, and (2) segmenting image to obtain the cow mask.

Step 1: We first subtract the two histograms, resulting in a cow histogram,$H_c(k)$:

$$H_c(k) = H_r(k) - H_b(k)$$  (6)

where $k$ is the depth value (in mm) in the range $1000 \leq k \leq 4000$. This range is restricted by the Kinect measurement range. Fig. 4e gives an example of the resulting cow histogram.

From $H_c$, we determine the minimum depth, $d_{\text{min}}$, and maximum depth, $d_{\text{max}}$, at which the cow appears in the depth image. These threshold values are determined by finding the longest sequence of distances in $H_c$ that holds positive values (see Fig. 4e), which is
\{d_{\text{min}}, d_{\text{min}+1}, \ldots, d_{\text{max}}, d_{\text{max}}\}. \text{ We then determine the mid-depth-value:} \]

\[ d = \frac{d_{\text{max}} + d_{\text{min}}}{2} \]  

**Step 2:** Now that the depth range in which the cow is positioned is determined, we can obtain the cow mask. To obtain this, we first calculate mask \( M_r \) containing all pixels in the restored depth image within a distance of \(-500\) to \(+1000\) mm from \( d \):

\[ M_r(p) = \begin{cases} 1 & \text{if } d - 500 \leq D_r(p) \leq d + 1000 \\ 0 & \text{otherwise} \end{cases} \]  

And we do the same for the background depth-image:

\[ M_b(p) = \begin{cases} 1 & \text{if } d - 500 \leq D_b(p) \leq d + 1000 \\ 0 & \text{otherwise} \end{cases} \]  

We then obtain the cow mask by subtracting these two masks:

\[ M_c(p) = M_r(p) - M_b(p) \]  

As this mask might contain some small bits of noise, we filter the mask using a median filter with a kernel of \( 5 \times 5 \):

\[ M'(p) = \text{median}(\{M_c(q)\mid q \in N^{5 \times 5}(p))\} \]
Fig. 3. Images before a and after b restoration.

Fig. 4. a. background image; b: restored image; c. Histogram of background; d: histogram of restored image; e, Hc.
2.2.3. Image selection

The following requirements are applied to ensure that the image contains a single cow and is well segmented. (1) the width and height of the object mask should be smaller than respectively 510 and 422, due to the image resolution of the Kinect v2 depth data of 512 × 424, (2) the size of the object, calculated as the number of pixels in the object mask, should be larger than 20,000, (3) the length–width ratio should be between 1.80 and 1.17 (cow’s side), 0.97 and 0.42 (cow’s back), and (4) object masks with holes that are larger than 500 pixels should be filtered to ignore incorrect cow-background segmentation close to the hooves connecting two separate legs.

2.2.4. Skeletonization

The detection and segmentation of the body parts is based on a skeleton description of the cow mask. To extract the cow skeleton, skeletonization was performed in 3 steps: (1) feature distance transform, (2) thinning, and (3) feature detection.

Step 1: feature distance transform

The feature distance transform (FDT) is adopted from Strzodka and Telea (2004) and is based on the standard distance transform (DT). The DT of a binary image is a well-known shape representation. Given an object mask with holes that are larger than 500 pixels should be filtered to ignore incorrect cow-background segmentation close to the hooves connecting two separate legs.

Step 2: thinning

Thinning is an image-processing operation in which binary image regions are reduced to curved lines with single-pixel thickness that approximate the center skeletons of the regions without breaking the objects apart (Lam et al. 1992). Thinning the initial skeleton from step 1 results in a one-pixel wide skeleton (Fig. 7c). The skeleton obtained this way is essentially a homotopic skeleton generated by a series of sequential morphological thinning using the structuring elements from the Golay alphabet (Golay, 1969) until convergence.

Step 3: skeleton-graph construction

In order to turn the skeletons of step 2 into skeleton graphs, we will consider the endpoints and bifurcation points of the skeleton. Endpoints are detected by applying a 3 × 3 sliding window and detect the number of skeleton pixels. If the center of the window is s skeleton pixel and the total number of skeleton pixels in the window is 2, an endpoint is detected. A similar approach can be used for bifurcations with the constraint that at least 4 pixels are active in the window. This will potentially give multiple detections. A decluttering step is used to interpolate between those points to approximate the true spatial position of the bifurcation. Then the skeletal branches are detected from endpoints to bifurcations based on an 8-connected-neighbors searching method. This results in a skeleton graph with the vertices consisting of the endpoints and bifurcations and the edges being the skeleton branches that connect these vertices (Fig. 7d). Shorter skeletal branches contain fewer skeleton features for template matching. As a final step, every branch length is calculated and the skeleton branches with a length shorter than 10 pixels are removed (Fig. 7c).

2.2.5. Skeleton classification

When we obtain the skeleton map, we need to classify each skeleton branch. The way to achieve this is to calculate the adjacent skeleton path and perform similarity matching with the templates in the database to obtain the best matching result. When we determine the category that each skeleton branch belongs to, we can divide the body into different parts.
2.2.5.1. Template matching. Template matching is based on a skeleton-graph-matching method via path similarity proposed by Bai (2008). The set of endpoints of the skeleton graph is denoted as \(\{e_1, e_2, \ldots, e_{N_0}\}\), where \(N_0\) is the number of endpoints. The set is ordered on the angle with respect to the center of gravity, \(G_c\), in an anti-clockwise direction, see Fig. 8. A skeleton path from endpoint \(e_m\) to \(e_n\) is denoted as \(p(e_m, e_n)\). \(G\) represents the skeleton of the cow for testing, and \(G'\) represents the skeleton of templates in the dataset. The shape dissimilarity between two paths, \(p(e_m, e_n)\) and \(p'(e'_m, e'_n)\), in two different skeleton graphs, \(G\) and \(G'\) is calculated using the path distance \(pd(p(e_m, e_n), p'(e'_m, e'_n))\). The path distance is calculated by sampling each path with \(M_0 = 40\) equidistant points, \(\{q_1, q_2, \ldots, q_{M_0}\}\) and \(\{q'_1, q'_2, \ldots, q'_{M_0}\}\), along the paths. The path distance is then calculated by:

\[
pd(p(e_m, e_n), p'(e'_m, e'_n)) = \sum_{i=1}^{M_0} \left( \frac{r_i - r'_i}{r_i + r'_i} \right)^2 + \frac{(l - l')^2}{l + l'} + \beta \frac{(\alpha - \alpha')^2}{\alpha + \alpha'}
\]

(15)

where \(r_i\) is the radius of the maximum inscribed circle at sample point \(q_i\), which can be sampled from the distance transform calculated in Section 3.2.4, \(l\) is the Euclidean distance of endpoint \(e_m\) and \(e_n\):

\[
l(e_m, e_n) = \sqrt{(e_{mx} - e_{nx})^2 + (e_{my} - e_{ny})^2}
\]

(16)

and \(\alpha\) is the angle of the line between the vector \(\overrightarrow{e_mG_c}\) from endpoint \(e_m\) and \(G_c\), and vector \(\overrightarrow{G_c e_n}\) between \(G_c\) and endpoint \(e_n\):

\[
\alpha(e_m, e_n) = \arccos\left( \frac{l(e_m, G_c) + l(e_n, G_c) - l(e_m, e_n)}{2l(e_m, G_c)l(e_n, G_c)} \right)
\]

(17)

The relative influence of the inscribed circles of the path sample points, the Euclidean distance between the end points and the angle between the end points can be weighted with parameters \(\delta\) and \(\beta\).

In our experiment, we sampled each skeleton path between two adjacent endpoints and compared with all other skeleton path. The reason is that we find there is a strong relationship within the \(r_i\), land \(\alpha\) between next two endpoints. But that means we need more templates.
Fig. 9. Samples of Database.
that can cover all the posture of dairy cow that introduced in 2.2.5.2.
Then, the matrix of the path distances is changed as, 

\[
\text{Dis}(e_i, e'_j) = \begin{pmatrix}
P_d(p(e_0, e_1), p(e'_0, e'_1)) & P_d(p(e_1, e_2), p(e'_1, e'_2)) & \cdots & P_d(p(e_N, e_1), p(e'_N, e'_1)) \\
P_d(p(e_0, e_1), p(e'_0, e'_1)) & P_d(p(e_1, e_2), p(e'_1, e'_2)) & \cdots & P_d(p(e_N, e_1), p(e'_N, e'_1)) \\
\vdots & \vdots & \ddots & \vdots \\
P_d(p(e_0, e_1), p(e'_0, e'_1)) & P_d(p(e_1, e_2), p(e'_1, e'_2)) & \cdots & P_d(p(e_N, e_1), p(e'_N, e'_1)) 
\end{pmatrix}
\]

(18)

The similarity calculation is to find out the minimum distance F with all skeleton templates in database,

\[
F = \min\{\text{Mean} + \text{Dev} + \sum_{i=0}^{k} \min\{\text{Dis}_{\text{sum}}(i)\}\}
\]

(19)

where Mean and Dev are stand for mean and standard deviation of \(\text{Dis}(i, i)\), respectively,

\[
\text{Mean} = \frac{\sum_{i=0}^{M} \text{Dis}(i, i)}{M_0}, \quad \text{Dev} = \sqrt{\frac{\sum_{i=0}^{M} [\text{Dis}(i, i) - \text{Mean}]^2}{M_0}}
\]

(20)

2.2.5.2. Template database. To create a training set with cows used for the template matching, we deploy an imaging system to capture cow images at the same experimental dairy farm at June 19, 2017, independent from the test data (see Fig. 9). The acquired cow images are 8-bit grayscale images with a resolution of 512x424 pixels. The developed database includes 221 cow images and it can be divided into 3 endpoints (11 images, Fig. 9a), 4 endpoints (50 images, Fig. 9b), 5 endpoints (79 images, Fig. 9c), 6 endpoints (57 images, Fig. 9d), 7 endpoints (19 images, Fig. 9e) and 8 endpoints (5 images, Fig. 9f). The training data was manually annotated.

2.2.5.3. Recognition of right or left leg. Discriminant factors \(Q_{\text{bf}}\) were defined for distinguishing far and near leg endpoint when two endpoints of forelegs or hind legs branches were detected (Bo et al., 2014). \(b_i\) is the skeleton branch that has endpoint \(e_i\) on the pruned skeleton graph \(G\) and contains skeleton points represented by \(b_{ik}(k = 1, 2, \ldots, L)\) and the corresponding depth image value is \(I(x_{bi}, y_{bi})\). With \(e_i\) as the starting point, the skeleton points along the skeleton branch can build a subset \(\{b_{ik}, b_{i(L−1)}, \ldots, b_{i(L−p)}\}\). Discriminant factor for near or far of endpoint \(e_i\) is then defined as the average depth of this subset:

\[
Q_{\text{bf}} = \frac{1}{\sum_{k=0}^{L-p} I(x_{bi}, y_{bi})}
\]

when it is judged that two end points \(e_i\) and \(e_j\) of \(G\) satisfy \(Q_{\text{bf}i} < Q_{\text{bf}j}\), \(e_j\) is classified as part of the skeleton branch of the proximal cow limb and \(e_i\) is classified as part of the skeleton branch of distal cow limb. Otherwise, \(e_i\) corresponds to distal and \(e_j\) to proximal. Distal and proximal are then be used to distinguish the left and right limbs of the cow.

Cow skeleton branches can be divided into eight groups: head, left foreleg (LFL), right foreleg (RFL), left hind leg (LHL), right hind leg (RHL), torso, udder and tail. Fig. 10 show results of template matching that were automatically generated results. And the endpoint ‘body’ was added in endpoints classification from back to recognize pin bone of cow when cows are thin.

2.2.6. Body-part segmentation

2.2.6.1. Body-part segmentation. Cow whole body can be divided into nine parts: head, left foreleg (LFL), right foreleg (RFL), left hind leg (LHL), right hind leg (RHL), torso, udder, belly (rear) and tail. It is combining gradient information and seeded region growing into segmentation process.

Step 1. Edge detection

Firstly, each body region is covered using inscribed circles along its skeleton branch from endpoint to bifurcation in order of LHL, torso, LFL, RFL, head, RHL, tail (side-view) and torso, LHL, RHL, head, LFL, RFL, tail (rear-view). The obtained mask \(S_i(x, y)\) provides a rough segmentation of the body-part in the depth image. To obtain a more detailed

![Fig. 10. Result of template matching for cow side (a) and cow back (b).](image-url)
segmentation, a region-growing algorithm is used (Gonzalez and Woods, 2018), which is separated by the full cow segment and the other body parts, preventing overlap between body parts. To distinguish adhesion body regions (e.g. tail and hindlegs), Schar operator was used to compute gradient in \( x \) and \( y \) directions in entire depth image. In the images shown in Fig. 11, different body-part are marked automatically. The partial derivatives \( G_x(p) \) and \( G_y(p) \) of the image \( f(p) \) along horizontal and vertical directions. The GM of \( f(p) \) is defined as

\[
\sqrt{G_x(p)^2 + G_y(p)^2} = \sqrt{[h_x(p)f(p)]^2 + [h_y(p)f(p)]^2}
\]

where \( T_G \) stands for depth threshold, in this experiment \( T_G = 5 \).

Step 2. Automatic seeded region growing process

Seeded region growing is a procedure that groups pixels or sub regions into larger regions based on predefined criteria and it requires seeds as additional input. The basic approach is to start with a set of seed points and grow the regions by appending to each seed’s neighboring pixels that have similar properties to the seed. The seeded region growing algorithm applied in this study is summarized as follows:

Initially, the region is composed of a single seed point \( R_0 \) that determined by the endpoints of each skeleton branch. \( R(j = 0, \ldots, 6) \) represents the growing of the different regions. Then check 4-nbitherhood pixels around the seed to determine whether they are similar to the seed one by one. If so, push the pixel into \( R_j \), otherwise...
disregard it. Each region is thus allowed to grow according to the following equation,

\[
R'_j = \begin{cases} 
R_{i-1} \cup P_{si} & \text{if } G(P_{si}) = 0 \\
R_{i-1} & \text{otherwise}
\end{cases}
\]

(23)

2.2.6.2. Udder segmentation. The program initially calculates the radius and center position information of the potential udder circle and then converts RGB potential udder region to HSV space. The H channel is used to characterize the colors range for udder detection. Udder detection that can be divided into 2 steps: udder center and radius calculation, color filter.

Step 1: A 2D chessboard calibration method was applied to match color to depth image. GML C++ calibration toolbox v0.75 was adopted to calibrate the color camera and depth camera separately and then calculate intrinsic parameters, rotation matrixes and translation vectors both for color and depth cameras.

Step 2: Skeleton structure is adopted to locate potential udder center for cow’s side. \(P_u\) is calculated as the udder center located on the maximal inscribed circle (green circle in Fig. 11a) of the bifurcation \(P_j\) of skeleton branch of hip. \(R_T\) is the circle radius, indicated by the red circle.

![Fig. 13. Partially Correct Solution.](image-url)
(Fig. 12a), calculated by the formula analyzed by curve fitting (MATLAB 2016b) using database,
\[
R^2 = -5.289 \times 10^{-2} \cdot \frac{P_r}{10}^2 + 1.068 \cdot \frac{P_r}{10} - 4.965, (R^2 = 0.971)
\]
where \(P_r\) means the whole coverage of pixels for cow body. For back-view udder, the radius and center position are calculated by equation (26) in the rear region described as grey area shown in Fig. 11b. center view udder, the radius and center position are calculated by equation (26).

\[
\begin{align*}
Q_x &= \frac{1}{2}(X_{r_{max}} + X_{r_{min}}), \\
Q_y &= \frac{1}{3}(2Y_{r_{max}} + Y_{r_{min}}) \\
R_L &= \frac{1}{3} \text{abs}(Y_{r_{max}} - Y_{r_{min}}) \\
R_S &= \frac{1}{3} \text{abs}(X_{r_{max}} - X_{r_{min}})
\end{align*}
\]

where \(X_{r_{max}} \) and \(X_{r_{min}}\) present maximal and minimal \(x\) value of edge points of rear zone, respectively. \(Y_{r_{max}}\) and \(Y_{r_{min}}\) present maximal and minimal \(y\) value of edge points of rear zone, respectively.

**Step 3:** the potential udder region in RGB was converted to HSV color space because it is more related to human color perception. The red color in channel \(H\) is characterized by values within \((0, 20)\|\((260, 360).\) The red circles in the black background (Fig. 13a&c) show the results of color filtering that udder region is enhanced. Then, a 3×3 median filter was used to soften more the results achieved by the dilatation and erosion. Color images shown in Fig. 12c&d represent the result of full body detection from side and back, respectively.

### 2.3. Experimental setup

The performance of body-part detection was evaluated using accuracy of endpoints classification and body-part segmentation.

#### 2.3.1. Skeleton classification

In this experiment, we evaluated the performance of the skeleton-classification method. We selected 1421 side-view images taken from 113 cows and 859 rear-view images from 75 cows at random. The performance is evaluated by presenting the results in a normalized confusion matrix, where the correct classifications and misclassifications are shown. Whether a body part was correctly classified by the skeleton-classification method was determined by 3 experts. In the case of discordance among experts we used majority voting.

The overall performance is measured using the classification rate (CR)

\[
CR = \frac{N_c}{N_r}
\]

where \(N_c\) is the numbers of correctly classified images and \(N_r\) is the number of images in total. The skeleton in an image is correctly classified when the skeletal branches were correctly classified.

To further analyze the performance, we classify images into three classes:

- **Correct Solution:** where all the branches appear visually correct.
- **Partially Correct Solution:** where at least a half of the branches appears visually correct.
- **Mostly Wrong Solution:** where visually the branches are perceived as wrongly detected.

#### 2.3.2. Body-part segmentation

In this experiment, we want to examine the accuracy of the proposed body-part segmentation method. A dataset containing 100 side-view images and 100 rear-view images taken from 50 cows was used to evaluate the performance. All images were manually labelled to get the ground-truth segmentation (GT). Fig. 12 shows an example of manual segmentation both from side and rear view. For each image in this dataset, the body-part segments resulting from the method were compared to the corresponding GT labelling. The quality of segmentation was evaluated using a pixel-level accuracy, sensitivity and specificity:

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}
\]

\[
\text{Sensitivity} = \frac{TP}{TP + FN}
\]

\[
\text{Specificity} = \frac{TN}{TN + FP}
\]

These parameters are analyzed and tabulated, where

- **TP** represents true positive: The number of body-part pixels in M correctly identified as body-part according to GT.
- **FP** indicates false positive: The number of body-part pixels in M that are non-body part in GT.
- **TN** indicates true negative: The number of non-body-part pixels in M correctly identified as non-body-part according to GT.
- **FN** indicates false negative: The number of non-body-part pixels in M that are body-part pixels in GT.

### 3. Results

#### 3.1. Skeleton classification

Tables 1 & 2 show the results of the skeleton-classification experiment for side and back views respectively. The tables show the normalized confusion matrices. The rows represent the true body parts and the columns represent the output of the method. The skeleton-classification method outputs eight classes: head, torso, left front leg (LFL), right front leg (RFL), left hind leg (LHL), right hind leg (RHL), hip and tail. “Others” means wrong skeleton branch which disappear on the body-part. Unidentified skeleton branches were classified as “Missing”. The classification rate for all parts is higher than 88.94%.

The results are presented in Tables 1 & 2 in the form of confusion matrix. The diagonal elements in the confusion matrix (Tables 1 & 2) shows the number of correctly classified instances. The bold values in table 1 and table 2 indicate correct rate. In the first row, the first element shows the proportion of skeleton branches belonging to ‘head’ class and classified as ‘head’. The second element shows the proportion of skeleton branches belonging to ‘head’ class but misclassified as ‘torso’. The third element shows the proportion of skeleton branches misclassified as LFL (left front leg) and so on.

The tables show what in general, the skeleton classification is successful. For most body parts, the classification rate is around 94−97%. Classification of the tail has a bit lower classification rate of 89% in case of the back views and 91% for the side view. Reason is that the tail is a thin structure that regularly is not visible in the cow silhouettes. Moreover, the tail often swings, which can result in the tail being filtered out by the temporal averaging filter in the image-restoration step.

For side-view cows, only one or two skeletal branches were wrongly classified. For instance, the “RHL” was classified as tail (Fig. 13a) and “tail” was classified as others (Fig. 13b). In all rear-view images, the wrong solution can be divided into partially correct solution which have only one wrong endpoint (43.37%) or with less than or equal to half of skeletal branches wrong (56.63%). For instance, the “Head” was classified as part of body (Fig. 13c) and “body” was classified as tail (Fig. 13d). The wrong classification with LHL and tail shown in Fig. 13a caused incorrect segmentation of LHL, udder and tail. The wrong classification with tail shown in Fig. 13b caused wrong segmentation of LHL.

Although Table 2 shows also high correct classifications for the back
from the side using depth images. We noted that the recognition rate of (2017), that employs RF to recognized eight body parts of dairy cow head, fore legs and tail have lower sensitivity because the rear-view torso. For cow right solution that hind legs, udder or tail were recognized as a part of torso is lower than that of Zhao study. Most of the errors involves the side and rear. we compare our system against the proposal of Zhao et al., the sensitivities were 70.58% and 44.93%, and the specificities were 97.05% and 95.07%.

3.2. Body-part segmentation

The representation of average accuracy, sensitivity and specificity are shown in Tables 3 & 4. In the detection of the body-parts of side-view and rear-view dairy cow, the average accuracies were 95.98% and 90.67%, the sensitivities were 70.58% and 44.93%, and the specificities were 97.03% and 95.07%.

In terms of accuracy, the torso accuracy was the lowest both from side and rear. we compare our system against the proposal of Zhao et al. (2017), that employs RF to recognized eight body parts of dairy cow from the side using depth images. We noted that the recognition rate of torso is lower than that of Zhao study. Most of the errors involves the wrong solution that hind legs, udder or tail were recognized as a part of torso. For cow’s side, the tail sensitivity was the lowest. For cow’s rear, head, fore legs and tail have lower sensitivity because the rear-view views, classifying endpoints is somewhat more complicated when viewed from the back. Fig. 13d shows wrong skeleton classification from the view of cows’ back. Fig. 13d shows that the branches of head and tail were wrongly segmented, causing wrong segmentations of head and tail. It also shows the missing of RFL. Observing the results, we noted that “tail” and “head” are more frequently wrongly classified (e.g. body or forelegs) or missed by the method. Most of the errors involved the swing of the tail mainly when depth images are imprecise or noisy due to sensor inaccuracies caused by direct sunlight from roof gap.

4. Discussion and conclusion

We have shown how for the first-time cow body-part detection for both of cows’ side and rear. The 3D depth and color images are acquired by the Microsoft Kinect v2 sensor and used with the distance transform values to effectively classify the cow body-parts. We estimated cow body-parts by classifying the branches of skeleton. The segmentation of the body-parts of dairy cows were tested in this study, and the test results showed that:

### Table 1

Result of skeleton classification for cows side (%).

<table>
<thead>
<tr>
<th>Body-part</th>
<th>Head</th>
<th>Torso</th>
<th>LFL</th>
<th>RFL</th>
<th>LHL</th>
<th>RHL</th>
<th>Hip</th>
<th>Tail</th>
<th>Others</th>
<th>Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head</td>
<td>97.05</td>
<td>0.00</td>
<td>0.35</td>
<td>0.28</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.28</td>
<td>2.04</td>
</tr>
<tr>
<td>Torso</td>
<td>0.00</td>
<td>96.55</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>3.45</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>LFL</td>
<td>0.14</td>
<td>0.00</td>
<td>94.44</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>3.80</td>
<td>1.62</td>
<td></td>
</tr>
<tr>
<td>RFL</td>
<td>0.28</td>
<td>0.00</td>
<td>94.02</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>LHL</td>
<td>0.00</td>
<td>0.00</td>
<td>96.22</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>2.04</td>
<td></td>
</tr>
<tr>
<td>RHL</td>
<td>0.00</td>
<td>0.00</td>
<td>0.42</td>
<td>0.00</td>
<td>0.00</td>
<td>96.06</td>
<td>0.00</td>
<td>0.07</td>
<td>3.17</td>
<td></td>
</tr>
<tr>
<td>Hip</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>97.89</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Tail</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.63</td>
<td>0.99</td>
<td>90.92</td>
<td>1.13</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2

Result of skeleton classification for cows back (%).

<table>
<thead>
<tr>
<th>Body-part</th>
<th>Head</th>
<th>Torso</th>
<th>LFL</th>
<th>RFL</th>
<th>LHL</th>
<th>RHL</th>
<th>Hip</th>
<th>Tail</th>
<th>Others</th>
<th>Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head</td>
<td>94.76</td>
<td>1.40</td>
<td>0.35</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>3.49</td>
<td></td>
</tr>
<tr>
<td>Torso</td>
<td>0.00</td>
<td>95.10</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.47</td>
<td>1.40</td>
<td>3.03</td>
<td></td>
</tr>
<tr>
<td>LFL</td>
<td>0.00</td>
<td>0.00</td>
<td>93.84</td>
<td>0.70</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>5.12</td>
<td></td>
</tr>
<tr>
<td>RFL</td>
<td>0.00</td>
<td>0.00</td>
<td>95.33</td>
<td>0.47</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.23</td>
<td>3.50</td>
<td></td>
</tr>
<tr>
<td>LHL</td>
<td>0.00</td>
<td>0.00</td>
<td>97.36</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.68</td>
<td>1.40</td>
<td></td>
</tr>
<tr>
<td>RHL</td>
<td>0.00</td>
<td>0.00</td>
<td>0.47</td>
<td>0.00</td>
<td>0.58</td>
<td>94.29</td>
<td>0.00</td>
<td>0.35</td>
<td>2.91</td>
<td></td>
</tr>
<tr>
<td>Hip</td>
<td>0.00</td>
<td>1.63</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>94.30</td>
<td>1.16</td>
<td>6.44</td>
<td></td>
</tr>
<tr>
<td>Tail</td>
<td>0.00</td>
<td>0.58</td>
<td>0.00</td>
<td>0.58</td>
<td>0.47</td>
<td>1.16</td>
<td>1.63</td>
<td>88.94</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>

### Table 3

Body-part Detection Performance Parameters Values for cow’ side.

<table>
<thead>
<tr>
<th>Body-part</th>
<th>Head</th>
<th>Torso</th>
<th>Belly</th>
<th>LFL</th>
<th>RFL</th>
<th>LHL</th>
<th>RHL</th>
<th>Hip</th>
<th>Tail</th>
<th>Others</th>
<th>Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.9644</td>
<td>0.8854</td>
<td>0.9620</td>
<td>0.9824</td>
<td>0.9219</td>
<td>0.9790</td>
<td>0.9773</td>
<td>0.9771</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.7825</td>
<td>0.9120</td>
<td>0.7558</td>
<td>0.7825</td>
<td>0.5356</td>
<td>0.8108</td>
<td>0.7300</td>
<td>0.2268</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specificity</td>
<td>0.9821</td>
<td>0.8559</td>
<td>0.9810</td>
<td>0.9821</td>
<td>0.9952</td>
<td>0.9879</td>
<td>0.9952</td>
<td>0.9952</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 4

Body-part Detection Performance Parameters Values for cow’ back.

<table>
<thead>
<tr>
<th>Body-part</th>
<th>Head</th>
<th>Torso</th>
<th>Rear</th>
<th>LFL</th>
<th>RFL</th>
<th>LHL</th>
<th>RHL</th>
<th>Hip</th>
<th>Tail</th>
<th>Others</th>
<th>Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.9638</td>
<td>0.6564</td>
<td>0.8611</td>
<td>0.9701</td>
<td>0.9725</td>
<td>0.9531</td>
<td>0.9553</td>
<td>0.9233</td>
<td>0.9038</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.2409</td>
<td>0.6154</td>
<td>0.5965</td>
<td>0.2280</td>
<td>0.2109</td>
<td>0.6493</td>
<td>0.6486</td>
<td>0.7783</td>
<td>0.0754</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specificity</td>
<td>0.9959</td>
<td>0.7033</td>
<td>0.9535</td>
<td>0.9944</td>
<td>0.9970</td>
<td>0.9870</td>
<td>0.9875</td>
<td>0.9407</td>
<td>0.9966</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1. In the classification of the skeleton of dairy cow, due to the distance limit and sensitivity of the sensor in the image acquisition, body-parts produced in the pre-processing was lost partially, so the precisions from side were higher than those from rear. An average detection rate for side-view and rear-view cows were 95.39% and 94.24%, respectively. In the detection of the body-parts of side-view and rear-view dairy cow, the average accuracies were 95.98% and 90.67%, the sensitivities were 70.58% and 44.93%, and the specificities were 97.03% and 95.07%. The average computing time for per RGB-D images was 4.94 s (Alienware NA75E-7HL, 2.8 GHz processor, 16 GB memory).

2. The evaluation of the body cleanliness for dairy cow is foundation for the segmentation of the body parts of the cow. The cleanliness of the cows did not significantly affect the segmentation of the cow’s body parts. However, the results of body segmentation will directly affect the accuracy of dirt detection. As the assessment areas of cleanliness includes belly, udder, near hind leg, rear and lower legs. In the absence of head, torso, front legs and tail, the key parts of the cleanliness evaluation could be accurately segmented, the average accuracies were 96.16% and 92.32%, the sensitivities were 77.82% and 66.82%, and the specificities were 98.06% and 96.72%, indicating that the method had a higher accuracy and was suitable for accurate segmentation of the key parts of cows.

3. The practical application of body segmentation system aims to achieve a new and useful improvement as welfare indicators (cleanliness evaluation). For commercial local farm conditions, such as camera location, background lighting, and milking routine will all need to be considered. Further research should investigate the methods of cleanliness evaluation on selected body-parts, e.g. belly, udder, rear and legs. The target of our article is to achieve the body segmentation for lactating dairy cows because the assessment of the cleanliness of the cows is mainly aimed at the cows involved in milking. Different types of cows or calves have different bone structure, muscles and hair. The shape and outlines of cow will be changed, so a new template dataset is needed.

In this paper, we have seen that we can successfully segment the images of cows in separate body parts in the majority of cases. Compared to the body-parts segmented by using previous method (Zhao et al., 2017; Jiang et al., 2019), the body-part segmented in this study are more fine and robust for both side-view and rear view cows, which are essential to segmenting cleanliness for dairy cow. With new training data, the proposed method can also be used for segmentation of other animals (e.g. horse or pig).

The presented work forms the basis of an automatic cleanliness scoring system. A dirt-detection algorithm needs to be developed and combined with our proposed body-part-segmentation method to be able to assess the cleanliness of each body parts. Where the current work used manual selection of images of side/rear views of standing or walking cows, the future system needs to acquire such images automatically. This can be achieved by placing the camera at the exit of the milking parlor, where the cows pass one by one in a predefined pose, and where illumination can be fully controlled, to not behampered by disturbances caused by sunlight.

CRediT authorship contribution statement


Declaration of Competing Interest

The authors declared that there is no conflict of interest.

Acknowledgements

This work was supported by China Agriculture Research System (CARS-36) and China Scholarship Council (CSC). Special thanks to staff at Sifang Experimental Dairy Farm in Datong for their extensive help and input during data collection at the farm.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compag.2020.105897.

References


