3-Dimensional pose estimation to detect posture transition in freestall-housed dairy cows

Adrien Kroese,[1]* **Moudud Alam,**[2] **Elin Hernlund,**[3] **David Berthet,**[4] **Lena-Mari Tamminen,**[1] **Nils Fall,**[1]
and Niclas Högberg[1]

[1]Department of Clinical Sciences, Faculty of Veterinary Medicine and Animal Science, Swedish University of Agricultural Sciences, Uppsala, Sweden, 756 51
[2]School of Information and Engineering, Dalarna University, Borlänge, Sweden, 783 33
[3]Department of Anatomy, Physiology and Biochemistry, Faculty of Veterinary Medicine and Animal Science, Swedish University of Agricultural Sciences, Uppsala, Sweden, 756 51
[4]Sony Nordic, Lund, Sweden 221 88

ABSTRACT

Freestall comfort is reflected in various indicators, including the ability for dairy cattle to display unhindered posture transition movements in the cubicles. To ensure farm animal welfare, it is instrumental for the farm management to be able to continuously monitor occurrences of abnormal motions. Advances in computer vision have enabled accurate kinematic measurements in several fields, such as human, equine, and bovine biomechanics. An important step upstream to measuring displacement during posture transitions is determining that the behavior is accurately detected. In this study, we propose a framework for detecting lying-to-standing posture transitions from 3-dimensional (3D) pose estimation data. A multiview computer vision system recorded posture transitions between December 2021 and April 2022 in a Swedish stall housing 183 individual cows. The output data consisted of the 3D coordinates of specific anatomical landmarks. The sensitivity of posture transition detection was 88.2%, and precision reached 99.5%. In analyzing those transition movements, breakpoints detected the timestamp of onset of the rising motion, which was compared with that annotated by observers. Agreement between observers, measured by intraclass correlation, was 0.85 between 3 human observers and 0.81 when adding the automated detection. The intra-observer mean absolute difference in annotated timestamps ranged from 0.4 s to 0.7 s. The mean absolute difference between each observer and the automated detection ranged from 1.0 s to 1.3 s. We found a significant difference in annotated timestamps between all observer pairs, but not between the observers and the automated detection, leading to the conclusion that the automated detection does not introduce a distinct bias. We conclude that the model is able to accurately detect the phenomenon of interest and that it is equitable to an observer.

Key words: computer vision, animal welfare assessment, freestall cubicle, pose estimation

INTRODUCTION

All cubicles in a dairy barn are usually identical, but a natural variability exists both in animal size relative to the cubicle (Dirksen et al., 2020) and in individual motion patterns and locomotor activity (Shepley et al., 2020). A factor of stall comfort, which affects lesion prevalence and lying time, is the ease with which a cow is able to get up and down in the cubicle (Zambelis et al., 2019). Ease of movement during posture transition was highlighted as an evaluation criteria for stall quality in relation to cow comfort by Lidfors (1989), who noted that cows in cubicles were more regularly seen performing abnormal motions (such as sideways lunging or horse-like rising) than those on pasture. Ceballos et al. (2004) analyzed the kinematics of posture transitions and found that cows used less longitudinal space when rising in a cubicle than on an open pack. Given the evidence for the link between restrictive movements and signs of reduced welfare (Beaver et al., 2021), the quality of posture transitions is included as an indicator in welfare assessment schemes such as Welfare Quality (Blokhuis et al., 2013).

Assessing ease of posture transition per se, rather than through indirect signs of reduced comfort such as hock lesions (Dirksen et al., 2020) or reduced lying time (Shewbridge Carter et al., 2021), is more challenging, and practical objective methods are needed (Brouwers et al., 2023). Visual observations noting the occurrence of abnormal behaviors are commonplace in farm management and welfare assessment schemes. Alternatively,

ease of movement can be assessed quantitatively by measuring the displacement of anatomical landmarks throughout bouts of posture transition (Ceballos et al., 2004). Drawbacks exist for both approaches. The visual method relies on time-consuming, sporadic human observations. Although Zambelis et al. (2019) found excellent agreement between observers (kappa of 0.93 for getting-up movement ease), a degree of subjectivity always exists in visual scoring of animal movements (Chaplin and Munksgaard, 2001; Vasseur, 2017). The acquisition of 3-dimensional (**3D**) kinematics data by Ceballos et al. (2004) relied on fitting motion-capture reflectors on cows, requiring lengthy preparation and exposure of the equipment to damage. These limitations might be a reason behind the low sample size (n = 5 cows with at least 2 bouts per cow) in the latter study.

Considering the variability in cow sizes and kinematic profiles and the need for objective methods to assess ease of movement, we propose a framework to detect lying-to-standing (**LTS**) posture transitions from 3D pose estimation data. As a step in validating the potential of this method, the aim of this study was to measure the performance of a feature extractor in detecting the onset of LTS posture transitions compared with the human eye.

MATERIALS AND METHODS

The study presented here was approved by the ethical committee Uppsala djurförsöksetiska nämnd under approval 5.8.18-13069/2021. The 3 Rs in animal research were considered when using existing video material, previously and noninvasively collected.

Location and Animals

Recordings were obtained at the Swedish Livestock Research Centre's dairy barn (Uppsala, Sweden). The herd comprises Swedish Holstein and Swedish Red cattle housed indoors with access to pasture 120 d a year, between May and September. Video was recorded on 30 separate days (midnight to midnight), sampled for convenience, between December 8, 2021, and April 28, 2022. Because the barn is lit at all times, recordings were obtained at all times of day. An average of 51 cows were present simultaneously in the pen, with individuals being added and removed throughout, for a total of 183 different individuals having visited the pen during the study period. A total of 7 RGB cameras (G3 Bullet, Ubiquiti) were placed around an area approximately one-quarter of the pen, located closest to the sorting gate to the milking robot, and oriented toward the rows of cubicles so that all cubicles in the study ward, including forward lunge room defined as the 60 cm beyond the head rail, were visible
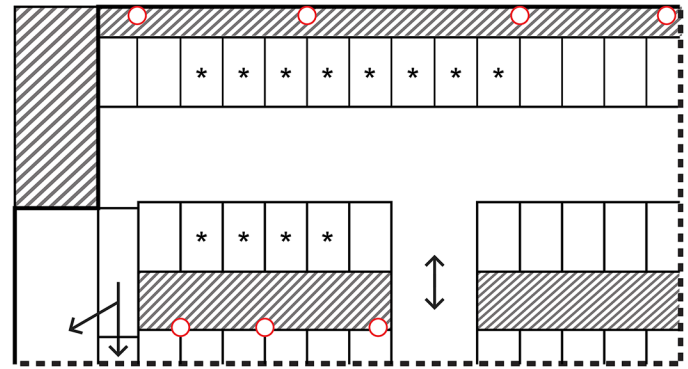

Figure 1. Schematic of the portion of the stall where recordings were obtained. The gray shaded areas are passageways unavailable to cows. Thick borders mark the stall boundaries, and dashed lines indicate a continuing area that is accessible to the cows beyond that shown here. Cameras are represented by red circles, placed between 2.8 and 3.6 m high. The parallel rectangles are cubicles; data were collected in cubicles marked with asterisks. The arrows indicate movement directions the cows are able to follow in the passageways.

by at least 2 cameras. The study ward comprised the 12 cubicles (CC1800 cubicle divider with rigid head bar, Delaval) for which video coverage was optimal, out of 66 total in the pen. The cameras were installed on fixed metal rails, part of the barn's infrastructure, between 2.8 and 3.6 m high. The locations of each camera, as well as the stall layout, are shown in Figure 1.

Cows had access to feeding troughs with ad-libitum mixed feed as well as 2 rotary brushes, and concentrate dispensed both at the milking robot and at concentrate dispensers. Passage through the milking robot's sorting gate was compulsory for access to the feed. Milking was done by one milking robot (VMS V300, Delaval), which cows had access to on a voluntary basis. Cows were brought to the robot by farm staff if they had not been milked in over 12 h.

Key Point Acquisition in 3 Dimensions

This study used 3D pose estimation software (Sony multi-camera system, Sony Nordic). The software estimates the 3D pose by finding cross-view correspondences across inferred 2-dimensional (**2D**) poses of the same object on synchronized views. It then creates a track for each object based on spatial continuity in the 3D location. The initial synchronization is achieved by reading the timestamp of each frame and relating the first full-second transition for a common timestamp across all video recordings as the initial synchronized frame. The initial frame synchronization is provided as an input to the multicamera system. Synchronization is maintained using the estimated time of arrival of each frame in the processing buffer.
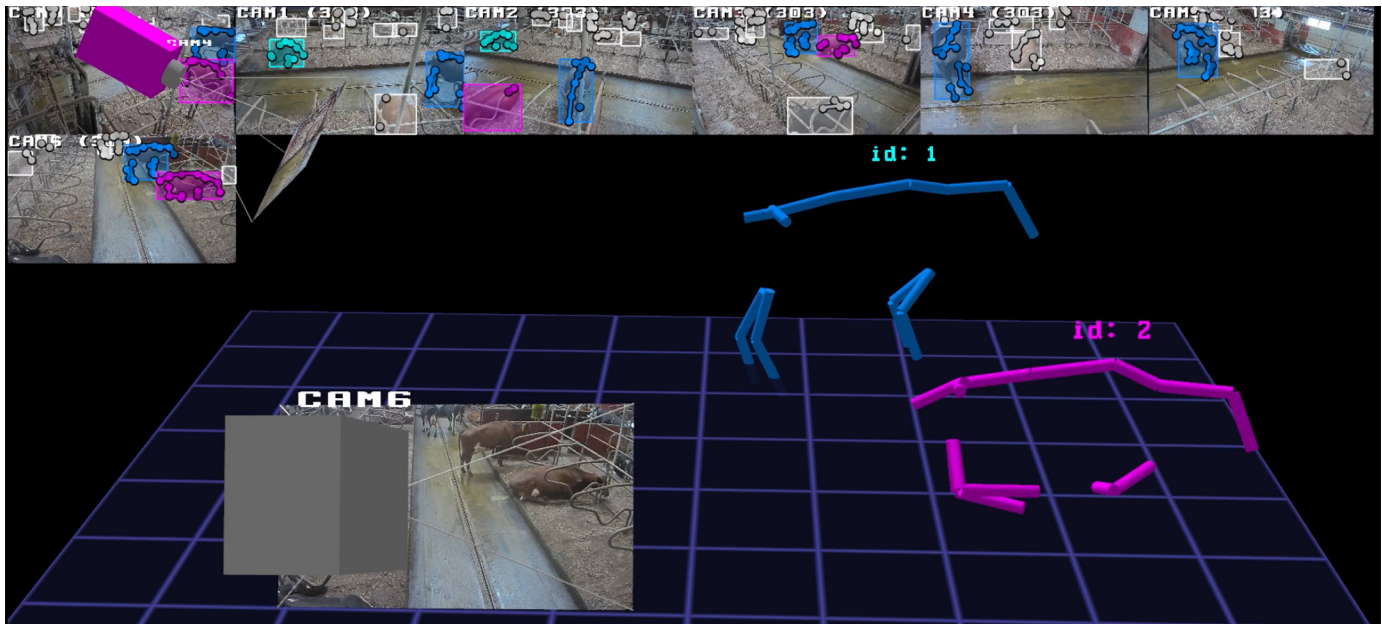
Figure 2. The 2D pose estimation and 3D fusion of 2 cows. The 2D results are displayed at the top, showing the synchronized frames from cameras 0 to 6, onto which predicted bounding boxes and key points are overlaid. The rest of the scene shows the projection of 2 cows from key points in 3D. Cameras 4 and 6 are represented as magenta and gray cuboids, respectively, in the 3D representation, in their spatial position relative to each other and to the cows. A projection of the frames from cameras 4 and 6 (identical to those in the 2D images above) is shown in front of the camera's 3D representation. The 5 other camera representations are not displayed from this angle, and camera 4 occludes the view from camera 0 because of the choice of angle. Only 4 of the key points shown in this figure were used in the study.

The 2D object detector and pose estimator use convolutional neural networks to detect cows and specific anatomical landmarks on RGB images, in the form of a bounding box and key points, respectively. The landmarks used in this study were limited to the center-top of the poll, the highest point at the withers, the spine at the 13th thoracic vertebra, and the top of the sacrum taken immediately behind the uppermost part of the ilium (referred to respectively as head, withers, t13, and sacrum).

The output data consists of one key point for each anatomical landmark with X, Y, and Z coordinates for each object and given frame. Figure 2 shows the estimated 3D position of the key points, linked to create a visual structure, for 2 objects during an LTS transition, as well as the video frames used to generate them.

Detection of Posture Transitions

The recordings were sampled visually by one observer with the aim of finding 1,000 sequences containing LTS transitions. When a cow was observed fully getting up from a lying positions, the timestamp was annotated, and a video sequence corresponding to a window of ± 15 s around the annotated timestamp was extracted. In the final data set, an arbitrary 979 sequences were eventually identified. These sequences were then processed with the 3D pose estimation software.

When the cow rises, the line formed by linking the sacrum and t13 key points increases its angle compared with the horizontal plane, as the cow's back is at an angle with the ground. By calculating the difference between the sacrum height and withers height, and following this difference through time, we identified peaks corresponding to LTS motions. When a peak above 0.4 (in the coordinates' arbitrary spatial reference system) was detected, the frame was considered to be within a potential rising motion. The mean withers Z position in the 120 frames located 330 frames after the peak was then compared with the mean withers Z position in the last 120 frames of the sequence. If the ratio of the height difference after and before the peak was higher than 140%, the track was classified as an LTS motion. Figure 3 illustrates this by showing the vertical position of the key points. At 16 s, there is an important difference in the heights of the withers (orange) and sacrum (green). This difference points toward a potential rising bout. Calculating the difference in withers position between the 5-s and 27-s marks, we determine that the animal has transitioned from a low, lying posture to a high, standing posture.

In these 979 sequences, this method initially detected 493 LTS motions for which the cow was tracked at each consecutive frame. For the remainder (486 sequences), the tracks were interrupted for several frames and the mo-
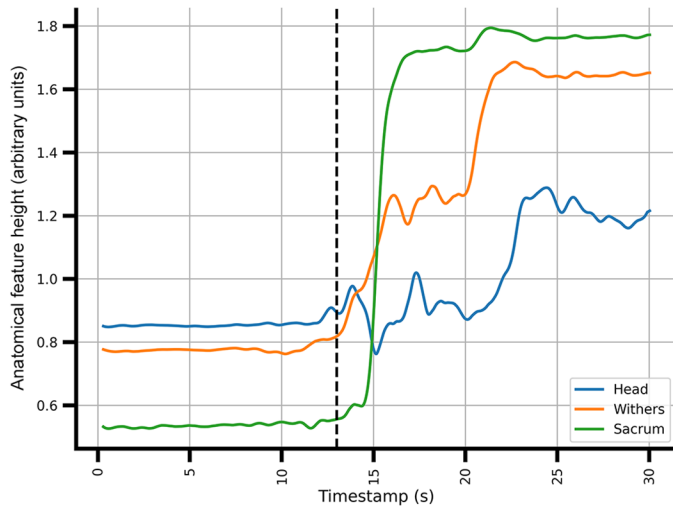
Figure 3. The coordinates of the anatomical landmarks of dairy cows were tracked with 3D pose estimation. This figure shows the Z coordinate (height) of a cow's head, withers, and sacrum throughout a lying-to-standing motion. Initially, the low variability on the vertical axis indicates that the cow is lying still. At about 11 s, the withers (orange) rise gently as the cow sits on its carps, followed by lunging with vertical bobbing of the head (blue) from 12 to 17 s. The sacrum (green) rises rapidly soon after, describing a sigmoid. There is a pause on the carps, with the sacrum already up, from 16 to 20 s. The cow has risen by the 22-s mark. The vertical dotted line shows the onset of the posture transition detected using linearly penalized segmentation. This example was selected for clarity.

tion was captured in several separate tracks. Detections were stitched together if they fit the following criteria:

- The tracks are found in the same 30-s sequence.
- The second track starts after the first track vanishes, and within an interval of 30 frames.
- The Euclidian distance in the 3D pose estimator's coordinate system between the last point in the vanishing track and the first point of the starting one is lower than 0.2.

No limit was imposed on the number of tracks appended together to form one single track, as long as the above conditions were fulfilled. The resulting stitched track was kept if it contained more than 700 frames, and discarded otherwise.

Using this method, an additional 370 rising sequences were detected by applying the height difference rule to the stitched tracks, giving a total of 863 predicted positives. For the remaining 116 sequences, either the animal was not detected by the pose estimation software, the posture transition detector failed to identify the occurrence, or the motion was split between different tracks that were not relatable due to noise or an interruption across more than 30 frames. Visual inspection of the predicted LTS

motions revealed 4 false positives. In addition, 22 true positives were discarded from the data set because the posture transition was initiated before the start of the video snippet and thus not captured in its entirety.

Signal Processing

Each series of raw coordinates was processed to attenuate noise. A low pass filter with a cutoff frequency of 10 Hz was applied to remove high-frequency noise resulting from key point jittering. This cutoff was chosen based on the recommendations by Hamäläinen et al. (2011) and Riaboff et al. (2020) for noise removal on animal activity data. The filter was applied separately to each key point and the respective time series of its X, Y, and Z coordinates. The filter was implemented in Python 3.9 (Python Software Foundation) using the function "butter" from the SciPy package (Virtanen et al., 2020). Figure 3 illustrates the filtered Z coordinates time series during a rising sequence.

From the processed signal, consisting of the coordinates of each key point in 3 dimensions, we detected the timestamp at which the cow starts rising. Considering solely the kinematic features available through the 4 key points, this is most clearly reflected by the change in the position of the withers, as rising on the elbows will cause the withers to rise upward slightly, which is visible by an increase in the withers' Z (vertical) coordinate. When doing so, the cow aligns its back along the length of the cubicle, which is reflected in a change of the withers' Y coordinate (axis perpendicular to the cubicle's length). Although, from a behavior perspective, there is more to the LTS transition than solely the withers' movement, the system was blind to all but the position of 4 anatomical landmarks. The withers were chosen for the stability of the key-point (low jittering) and for their consistent motion pattern in the LTS transition across sequences. To detect the exact onset of rising motions, we used linearly penalized segmentation (Pelt), implemented the Python library "Ruptures" (Truong et al., 2020). Pelt was applied to the bivariate series of the Y (lateral, perpendicular to the cubicles) and Z (height) positions of the withers to identify breakpoints in the time series. No restrictions were set on the number of breakpoints to be detected. A baseline height (Z coordinate) was calculated for each sequence as the median withers height in the first 30 frames of the sequence. The break points detected by Pelt were iterated through. If the median withers height in the 30 frames following the breakpoint was higher than the baseline, the breakpoint was then considered to be the start of the rising motion. If not, we iterated to the next breakpoint and applied the same logic.

Data processing, feature extraction, and analyses were carried out in Python 3.9.3 using the packages NumPy 1.21.5 (Harris et al., 2020) and SciPy 1.9.1 (Virtanen et al., 2020).

Validation Experiment

To evaluate the performance of the tool in detecting the occurrence of LTS bouts, we compared the timestamps automatically detected to those annotated by 3 human observers, considered as the gold standard for behavioral observations. Observers were provided with the following definition: "The cow is lying down and rises on its breastbone and elbows, which causes the withers to rise visibly above the rest of the back." This definition is based on that of Lidfors (1989), but it adds the position of the withers as an indicator. The animals were seen to initiate the movement by centering their elbows under the body, this in turn causes the withers to rise slightly. This motion of the withers was used to determine the exact onset of the rising motion. The description was accompanied by illustrations taken from Schnitzer (1971) and Cermak (1988), as well as an ethogram describing the sequence of movements in the LTS transition, in which the movement to label was explicitly identified. This ethogram described the stages of the posture transition based on Lidfors (1989) and on Schnitzer (1971). Observers all received the same training, in which the ethogram was explained and examples were showcased; they reviewed 5 videos of different cows rising and agreed on the exact frame to label as the onset of the rising motion. These 5 videos were taken from the original data set and used solely for training the observers.

The validation data set was sampled randomly from the 471 complete LTS sequences captured in a single track. In total, 60 unique LTS sequences were annotated by at least 1 observer. This number was determined a priori, as no prior data were available on observer variability in posture transition detection. These sequences were the original 30 s synchronized video snips from which the key points were detected. The video was available to the observers from all 7 cameras used for key point detection, plus one additional ceiling mounted camera. Observers were free to choose the camera offering the best view of the animal performing the bout. Every observer was provided with a total of 55 randomly selected video clips. Of these 55 sequences, 30 were common to all observers and 10 were unique to each observer (40 different sequences per observer). The remaining 15 sequences were randomly resampled from the prior 40 and re-annotated by the same observer, to measure intra-observer reliability. All sequences were blinded, with a different label each time the sequence appeared.

Statistical Analysis

The mean absolute difference (**MAD**) in annotate timestamp was calculated between each observer to quantify intra-observer reliability as $MAD(i) = \frac{1}{15}\sum_{s_i=1}^{15}\left|\Delta_{s_i}\right|$, where i = 1, 2, or 3 (observers) and $\Delta_{s_i} = t_{s_i,1} - t_{s_i,2}$, with $t_{s_i,1}$ and $t_{s_i,2}$ being the time stamp of the sth sequence provided at first and second assessment occasion, respectively, by observer i. In addition, the inter-rater MAD was calculated as $MAD(i,j) = \frac{1}{30}\sum_{s=1}^{15}\left|\Delta_{s,(i,j)}\right|$, where $\Delta_{s,(i,j)} = |t_{s,i} - t_{s,j}|, \forall i \neq j = 1,2,3$, with $t_{s,i}$ being the time stamp of the sth common sequence by ith observer. Mean differences (**MD**), as, for example, $MD(i) = \frac{1}{15}\sum_{s_i=1}^{15}\Delta_{s_i}$, were calculated, and the normality of Δ_{s_i} and $\Delta_{s,(i,j)}$ was assessed visually on a q-q plot. Subscripts i and j refer to 2 distinct observers: i = 1, 2, or 3; j = 2 or 3. The MD and MAD indicate interobserver systematic bias and dispersion, respectively.

The following mixed effects models were fitted using statsmodels.formula.api.mixedlm (Seabold and Perktold, 2010) in Python 3.9 to evaluate the observer effect and intraclass correlation (ICC) with or without the automated detection:

$$t_{s,i,r} = \beta_0 + \beta_1 I(i=2) + \beta_2 I(i=3) + u_s + \varepsilon_{s,i,r}, \quad [1]$$

$$t_{s,i,r} = \beta_0 + \beta_1 I(i=2) + \beta_2 I(i=3) + \beta_3 M + u_s + \varepsilon_{s,i,r}, \quad [2]$$

where β_0 is the (fixed) intercept, $u_s \sim N(0,\sigma_u^2)$ is a random sequence effect, s = 1 to 40 is the sequence indicator, β_1 and β_2 are fixed observer effects, β_3 is a fixed effect corresponding to the automated detection taken as an additional observer (referred to as the "model" or M), and $\varepsilon_{s,i,r} \sim N(0,\sigma_e^2)$ is a (random) error term. The sequence number is indicated by the subscript s, I_i are the observers, and r = 1, 2 is the index for repeated sequences annotated 1 to 2 times by the same observer. The observer effects were tested using ANOVA. The ICC as a measure of interobserver agreement were calculated as $ICC = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$. A post hoc pairwise t-test with Bonferroni correction for 6 tests was then computed to test the pairwise differences between observers. The annotated timestamps were not normalized because a 1 s difference between observers, for example, has the same practical

meaning in this context regardless of whether the annotation is done at the 4-s mark or the 12-s mark.

The performance of the algorithm was assessed in the same way, by treating the algorithm as an additional observer and seeing if it differed from the human observers. The differences were calculated between the algorithms' detection (denoted T^M) and the observer annotation, T^H. Bland-Altman plots were prepared for each observer pair $\left(T_i, T_j\right) = \left(\{t_{s,i}\}, \{t_{s,j}\}\right)$, and also comparing T^H with T^M, with a view to checking for the absence of a pattern and points beyond 1.96 standard deviations. Lastly, MAD(H, M), and MD(H, M) were calculated.

RESULTS

A total of 836 rising bouts were detected out of 979 visually selected sequences equating to a sensitivity of 88.5% or a false negative rate of 11.5%. Four sequences were wrongly classified as rising motions giving a precision of 99.5% or false positive rate of 0.5%.

Model 1, comparing only human observers, gave ICC = 0.85. We found a significant observer effect in predicting the annotated timestamps of LTS onset ($P < 0.001$) according to the ANOVA. When the model 2 was fitted to assess performance of the prediction, the ICC decreased to 0.81, remaining at a similarly satisfactory level of agreement. However, we found no significant difference between the predicted timestamp ("model") and each observer's annotations according to the post hoc pairwise t-test with Bonferroni correction of the type-1 error at α = 0.0083. We identified a significant difference between all observer pairs: $P\ (T_1, T_2) = 0.0016$; $P\ (T_1, T_3) < 0.001$; $P\ (T_1, T_3) = 0.0018$.

Mean absolute differences $T_s^{M,H}$ are summarized in Table 1. These values indicate good interobserver agreement and good agreement between humans and machine. The magnitude of $T_s^{M,H}$ is identical to that of $T_s^{M,M}$, meaning that T^M could be used in further research, as the model does not deviate from the observers more than they do from one another. Figure 5 shows the timestamp annotated by each observer (including the model and repeat sequences) for each sequence.

Intra-observer reliability was assessed using the mean absolute difference in seconds, and consistency using the standard deviation (σ). Observer 1 had an MAD of 0.55 $\pm$ 0.88 s ($\mu \pm \sigma$). Observer 2 had an MAD of 0.68 $\pm$ 1.47 s, and observer 3 had an MAD of 0.36 $\pm$ 0.48 s. The pooled standard error was 0.27 s. The standard deviation is preferred here to the standard error to quantify the variability in the differences between and within observers in annotated timestamps, independently of the number of samples. These results indicate very good intra-observer reliability (under 1 s on average).

Finally, we compared the annotations to the automated detections visually using the Bland-Altman plot in Figure 4. The upper left plot shows most points to be centered around 0, without signs of consistent bias from the model. More importantly, the spread was similar when comparing observers to the algorithm and observers together.

DISCUSSION

The ICC values show a good agreement between automated model detection and human observers in detecting the onset of cows' rising motions, according to previous research on the use of ICC as a reliability metric in animal motion scoring (Kaler et al., 2009). The ANOVA demonstrated a significant observer effect, strengthening the claim that observations of cows' movements are prone to individual variations. The post hoc test showed a significant difference in annotated timestamps between all pairs of observers, but the difference between the model and the observers was not significant. We conclude from this that the model's detection lies somewhere in between the observers' annotations. The MD of −0.06 s between observers and the model (Figure 4) and the proximity of the points to 0 show that no systematic bias was introduced by the automated detection. This latter finding is also supported by Figure 5, showing the timestamp annotated by each observer at each sequence, in which there is no evidence of the detection being consistently divergent from human annotations, as the triangular points (model) are not systematically above or below the circular ones (observers). We also see that the predictions do not tend to be further from the annotations than the annotations are from each other.

This agreement is a crucial step in validating the capability of 3D computer vision to accurately identify this specific kinematic feature in bovine behavior. Notably, the findings suggest that the model's performance does not considerably differ from human observers when compared with the variability among human observers. This suggests that the model does not introduce a distinct source of error in the detection process. Although discrepancies exist between the model and human observations, the magnitude of these divergences is not meaningful in comparison to the overall duration of the LTS transition.

Table 1. Interobserver agreement (MAD $\pm$ σ) between the annotations of all pairs of observers, including the model; pairs between observers calculate the MAD on 30 sequences, whereas pairs with the model include an additional 10 annotations, unique to each observer

Item	Observer 1	Observer 2	Observer 3
Model	1.02 ± 1.41	1.00 ± 1.70	1.30 ± 1.45
Observer 1		1.10 ± 1.26	1.67 ± 1.72
Observer 2			0.89 ± 1.01
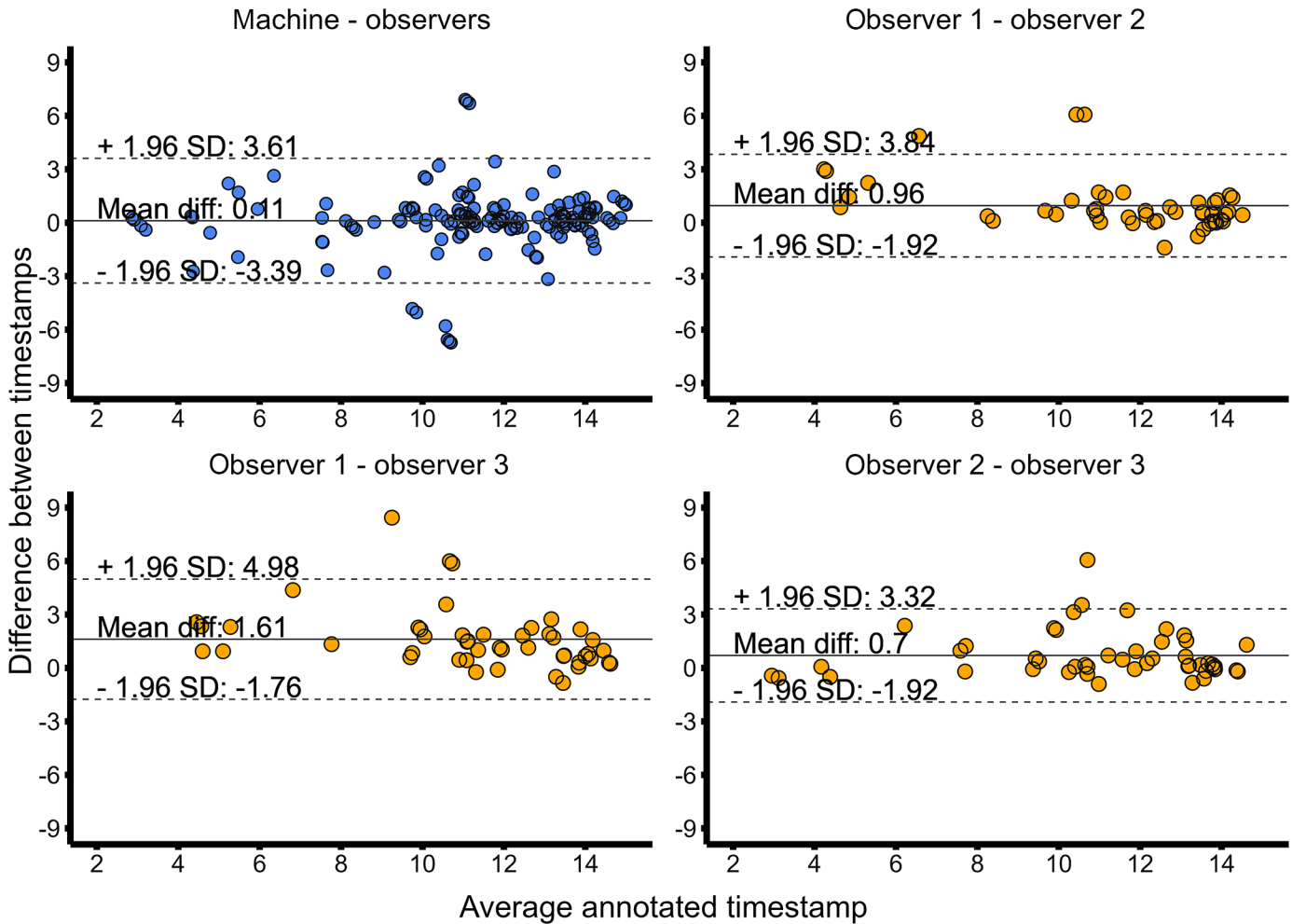
Figure 4. Bland-Altman plots comparing the timestamp of onset of cows' rising motions annotated by human observers to that predicted by the model. The 3D pose estimation provided the coordinates of cows' anatomical landmarks. Detecting breakpoints in the key point motion enabled detection of the onset of rising. Diff = difference. All units in seconds.

However, some limitations are important to mention. One such limitation is the likely over-representation of specific individuals. The animals were filmed in a limited area of the barn, and we can expect a degree of site fidelity from the animal (Vázquez Diosdado et al., 2018), leading to some individuals being over-represented. Because there was no individual detection, correcting for individuals was not possible. It is also unlikely that all recorded bouts were spontaneous; some may have been triggered by human intervention or by the presence of agonistic individuals. Bout motivation could introduce changes in kinematic patterns and velocity and potentially affect the accuracy of the automated detection.

Limitations also exist regarding external validity, as the study was conducted in a single cubicle design, under a limited period of time, and using manually selected video sequences. This manual selection work upstream of the automated processing is an important limitation

that drove the high sensitivity and specificity. The same system should be tested on continuous recordings. To counterbalance this limitation, however, the posture transition is an evident behavior, with a large difference in key point height before and after, which would easily be captured even with noisy key points by simply following the height of the cow's back.

The scope of this study was determined retrospectively; the decision to compare the automated detection to manual annotations was made after collecting the data and visually identifying LTS motions. The inclusion criteria were based on data quality and not experimental considerations. The exclusion of 22 longer bouts discarded important information with implications for the most vulnerable individuals when it comes to stall comfort, as a long pause during the posture transition is associated with adverse welfare outcomes (Zambelis et al., 2019)
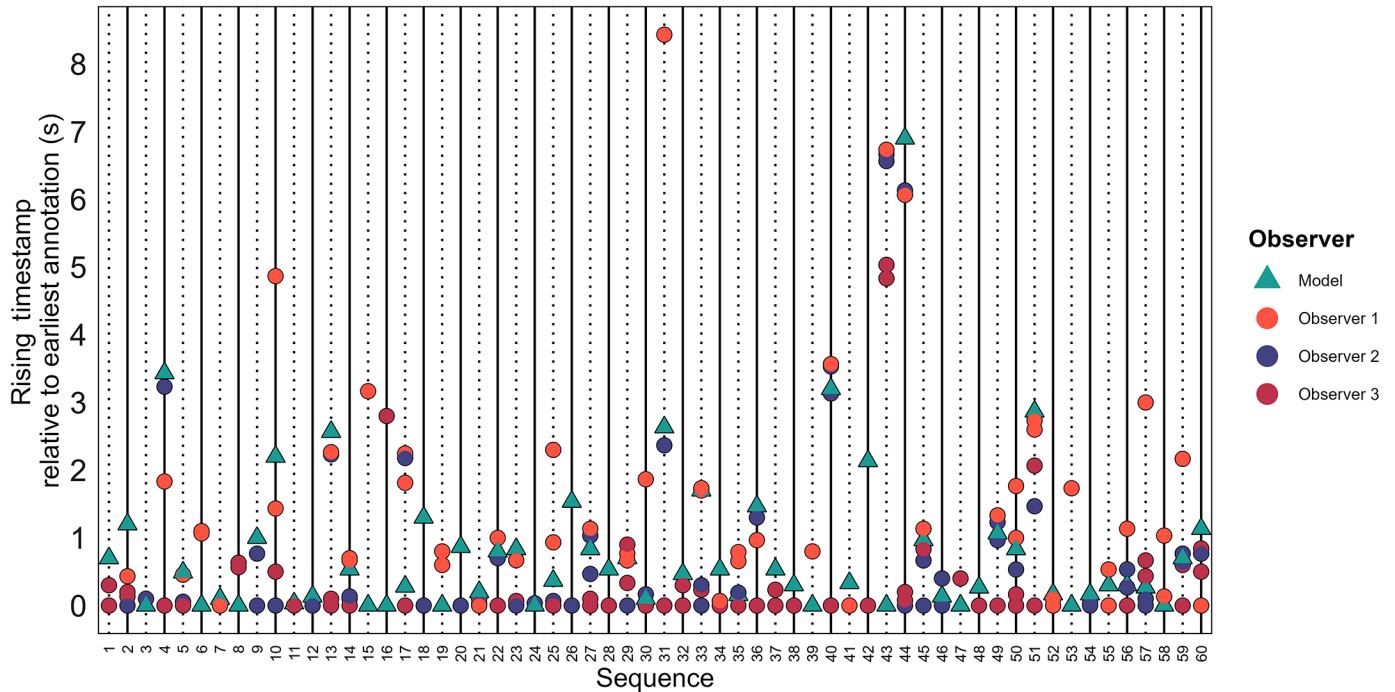
Figure 5. Annotated timestamp by each observer and by the model. The discrete x-axis shows each lying-to-standing sequences. On the y-axis is the timestamp of the onset of the posture transition annotated by each observer or predicted. From each annotation is subtracted the earliest timestamp in that sequence.

The study's gold standard was human observation, which is known to be variable across observers due to individual subjectivity. Although a bias is incorporated in the model, this bias is consistent across observations. The accuracy of the model could be improved by both altering the ethograms to make them more "machine-learnable" (Brouwers et al., 2023) and by diversifying the data. Importantly, although human observations are biased, humans are rarely completely incorrect, especially when the phenomenon at hand, such as posture transition, is evident. Algorithms on the other hand sometimes produce unexpected results, and monitoring and understanding their occurrence is essential for practical application. For instance, a difference of 6 s is found between the model and observer 2 in sequence 31 (Figure 5). Upon visual inspection of this sequence, the algorithm picked up on the onset of the adjustment movements, which were particularly lengthy in this sequence, making up the initial part of the posture transition. The second observer, on the other hand, noted the moment the fast rising motion occurred. This is not an error of either method, but a misalignment in the interpretation of the behavior. Referring to the description of the behavior provided to the observers, and quoted in the Materials and Methods section, the timestamp

automatically detected is closer to the phenomenon of interest.

Most significant for this research is that automated detection via computer vision offers an objective method for detecting specific motions, which is desirable for studies of behavior and motion patterns. Judging by the advances in equine kinematic research, markerless computer vision constitutes both a robust and practical data acquisition tool to measure the displacement of anatomical landmarks, offering similar accuracy to motion capture, albeit for specific motions (Lawin et al., 2023). Reliably identifying the motion of interest is only a step in the study of posture transition kinematics, which contain welfare indicators (Zambelis et al., 2019), the measure of which can be automated (Brouwers et al., 2023). Future studies using this technology aim at implementing individual recognition, which could contribute to a pool of sensor data at individual level. However, in the absence of individual identification, this technology is still able to deliver meaningful information either at herd or at cubicle level. The automated detection through 3D computer vision could, after further validation, serve as a new gold standard for the task of detecting LTS transitions (and other movements), similar to how interpreting accelerometer

data has become standard in behavior classification of ruminants (Riaboff et al., 2022).

CONCLUSIONS

In summary, our results demonstrate good agreement between human observations and automated detection of cows' rising motions. Notably, they indicate that the model introduces no more bias than human observers. This finding validates the use of multiview 3D pose estimation for detecting the onset of rising motions in bovine behavior, albeit in the conditions of a single farm. Automating the task with computer vision presents an opportunity to scale up bovine kinematic measurements and behavior monitoring and apply objective methods to further study.

NOTES

Nonstandard abbreviations used: 2D = 2-dimensional; 3D = 3-dimensional; diff = difference; ICC = intraclass correlation; LTS = lying-to-standing; MAD = mean absolute difference; MD = mean difference.

REFERENCES

Beaver, A., D. M. Weary, and M. A. G. von Keyserlingk. 2021. Invited review: The welfare of dairy cattle housed in tiestalls compared to less-restrictive housing types:—A systematic review. J. Dairy Sci. 104:9383–9417. https://doi.org/10.3168/jds.2020-19609.

Blokhuis, H., M. Miele, I. Veissier, and B. Jones. 2013. Improving Farm Animal Welfare: Science and Society Working Together—The Welfare Quality Approach. Wageningen Academic Publishers.

Brouwers, S. P., M. Simmler, P. Savary, and M. F. Scriba. 2023. Towards a novel method for detecting atypical lying down and standing up behaviors in dairy cows using accelerometers and machine learning. Smart Agric. Technol. 4:100199. https://doi.org/10.1016/j.atech.2023.100199.

Ceballos, A., D. Sanderson, J. Rushen, and D. M. Weary. 2004. Improving stall design: Use of 3-d kinematics to measure space use by dairy cows when lying down. J. Dairy Sci. 87:2042–2050. https://doi.org/10.3168/jds.S0022-0302(04)70022-3.

Cermak, J. 1988. Cow comfort and lameness: Design of cubicles. Bov. Pract. (Stillwater) 23: 79–83. https://doi.org/10.21423/bovine-vol0no23p79-83.

Chaplin, S., and L. Munksgaard. 2001. Evaluation of a simple method for assessment of rising behaviour in tethered dairy cows. Anim. Sci. 72:191–197. https://doi.org/10.1017/S1357729800055685.

Dirksen, N., L. Gygax, I. Traulsen, B. Wechsler, and J.-B. Burla. 2020. Body size in relation to cubicle dimensions affects lying behavior and joint lesions in dairy cows. J. Dairy Sci. 103:9407–9417. https://doi.org/10.3168/jds.2019-16464.

Hamäläinen, W., M. Järvinen, P. Martiskainen, and J. Mononen. 2011. Jerk-based feature extraction for robust activity recognition from acceleration data. 11th Int. Conf. on Intelligent Systems Design and Applications. IEEE. https://doi.org/10.1109/ISDA.2011.6121760.

Harris, C. R., K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant. 2020. Array programming with NumPy. Nature 585:357–362. https://doi.org/10.1038/s41586-020-2649-2.

Kaler, J., G. J. Wassink, and L. E. Green. 2009. The inter- and intra-observer reliability of a locomotion scoring scale for sheep. Vet. J. 180:189–194. https://doi.org/10.1016/j.tvjl.2007.12.028.

Lawin, F. J., A. Byström, C. Roepstorff, M. Rhodin, M. Almlöf, M. Silva, P. H. Andersen, H. Kjellström, and E. Hernlund. 2023. Is markerless more or less? Comparing a smartphone computer vision method for equine lameness assessment to multi-camera motion capture. Animals (Basel) 13:390. https://doi.org/10.3390/ani13030390.

Lidfors, L. 1989. The use of getting up and lying down movements in the evaluation of cattle environments. Vet. Res. Commun. 13:307–324. https://doi.org/10.1007/BF00420838.

Riaboff, L., S. Poggi, A. Madouasse, S. Couvreur, S. Aubin, N. Bédère, E. Goumand, A. Chauvin, and G. Plantier. 2020. Development of a methodological framework for a robust prediction of the main behaviours of dairy cows using a combination of machine learning algorithms on accelerometer data. Comput. Electron. Agric. 169:105179. https://doi.org/10.1016/j.compag.2019.105179.

Riaboff, L., L. Shalloo, A. F. Smeaton, S. Couvreur, A. Madouasse, and M. T. Keane. 2022. Predicting livestock behaviour using accelerometers: A systematic review of processing techniques for ruminant behaviour prediction from raw accelerometer data. Comput. Electron. Agric. 192:106610. https://doi.org/10.1016/j.compag.2021.106610.

Schnitzer, U. 1971. Abliegen, Liegestellungen und Aufstehen beim Rind im Hinblick auf die Entwicklung von Stalleinrichtungen für Milchvieh. Kuratorium Für Technik Und Bauwesen in Der Landwirtschaft-Bauschriften 10:43.

Seabold, S., and J. Perktold. 2010. Statsmodels: Econometric and Statistical Modeling with Python. Pages 92–96 in Proc. 9th Python in Science Conference.

Shepley, E., J. Lensink, and E. Vasseur. 2020. Cow in motion: A review of the impact of housing systems on movement opportunity of dairy cows and implications on locomotor activity. Appl. Anim. Behav. Sci. 230:105026. https://doi.org/10.1016/j.applanim.2020.105026.

Shewbridge Carter, L., S. M. Rutter, D. Ball, J. Gibbons, and M. J. Haskell. 2021. Dairy cow trade-off preference for 2 different lying qualities: Lying surface and lying space. J. Dairy Sci. 104:862–873. https://doi.org/10.3168/jds.2020-18781.

Truong, C., L. Oudre, and N. Vayatis. 2020. Selective review of offline change point detection methods. Signal Processing 167:107299. https://doi.org/10.1016/j.sigpro.2019.107299.

Vasseur, E. 2017. Animal Behavior and Well-Being Symposium: Optimizing outcome measures of welfare in dairy cattle assessment. J. Anim. Sci. 95:1365–1371. https://doi.org/10.2527/jas.2016.0880.

Vázquez Diosdado, J. A., Z. E. Barker, H. R. Hodges, J. R. Amory, D. P. Croft, N. J. Bell, and E. A. Codling. 2018. Space-use patterns highlight behavioural differences linked to lameness, parity, and days in milk in barn-housed dairy cows. PLoS One 13:e0208424. https://doi.org/10.1371/journal.pone.0208424.

Virtanen, P., R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, and P. van Mulbregt., and SciPy 1.0 Contributors. 2020. SciPy 1.0: Fundamental algorithms for scientific computing in Python. Nat. Methods 17:261–272. https://doi.org/10.1038/s41592-019-0686-2.

Zambelis, A., M. Gagnon-Barbin, J. St John, and E. Vasseur. 2019. Development of scoring systems for abnormal rising and lying down by dairy cattle, and their relationship with other welfare outcome measures. Appl. Anim. Behav. Sci. 220:104858. https://doi.org/10.1016/j.applanim.2019.104858.

ORCIDS

Adrien Kroese ⦿ https://orcid.org/0009-0001-5780-7345
Moudud Alam ⦿ https://orcid.org/0000-0002-3183-3756
Elin Hernlund ⦿ https://orcid.org/0000-0002-5769-3958
David Berthet ⦿ https://orcid.org/0009-0006-3106-0622
Lena-Mari Tamminen ⦿ https://orcid.org/0000-0001-6781-4533
Nils Fall ⦿ https://orcid.org/0000-0001-5597-2358
Niclas Högberg ⦿ https://orcid.org/0000-0002-2672-7924