133. Video-based analysis of dairy cow behaviour: detection of lying down and standing up

I. Adriaens*, W. Ouweltjes, I. Hulsegge and C. Kamphuis

Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; ines.adriaens@wur.nl

Abstract

Digital agriculture offers opportunities for improved monitoring and precision phenotyping of farm animals, crucial to achieving a more sustainable livestock production sector. Video-based analysis enables the quantification of animal behaviour in a non-invasive, automated way with few sensors. To unlock its full potential, appropriate computer vision techniques are needed. In this study, we propose an algorithm to detect lying-down and standing-up behaviour in dairy cows based on changes in bounding box properties detected via YOLOv5 and tracked with DeepSORT. We analysed 86 videos with a standing-up or lying-down event. With different criteria applied to the bounding box time series, we could detect up to respectively 92.3 and 80% for standing-up and lying-down events, respectively, with an accuracy of less than 2 seconds. Using bounding box properties as proxy for body shape and location, a general cow detection algorithm can serve multiple behavioural analyses simultaneously, whilst interpretability of the algorithms is maintained.

Introduction

Today's society faces a plethora of challenges to achieve a sustainable, socially accepted, safe and sufficient food production. Digital agriculture and related technological solutions contribute to addressing these challenges, as they allow to monitor and phenotype production entities in an automated and standardized way (Basso and Antle, 2020). In precision livestock farming, both individual animals and groups are monitored via automated sensor technologies, providing farmers, veterinarians, nutritionists and other stakeholders with information on health, fertility and welfare. Besides individual on-animal sensors, more and more also off-animal technologies that allow simultaneous and non-invasive monitoring of multiple animals emerge.

A main challenge in precision livestock farming is the translation of raw and sometimes unstructured sensor data into valuable animal or herd parameters that can be used for decision support. Often, this translation needs dedicated tools for data interpretation, including animal identification, feature extraction, time-series modelling and detection of abnormalities. The development of robust and targeted data-processing pipelines is therefore a crucial step in the digital agricultural transition (Basso and Antle, 2020).

Automated analysis of animal behavior provides new tools for welfare and health monitoring and (precision) phenotyping, as many links between health disturbances and behavior in livestock are known. Traditionally, behavior is observed visually, rendering it a subjective and tedious task with only limited value in commercial farm settings. Automation of behavior assessment allows to use behavioral traits in monitoring and decision support tools for the entire herd with limited observation costs, and is therefore seen as an important focus in current research and development (Rushen *et al.*, 2012).

R.F. Veerkamp and Y. de Haas (eds)

DOI: 10.3920/978-90-8686-940-4_133, © I. Adriaens *et al.* 2022

In dairy cows, lying behavior changes upon disease such as lameness, mastitis and metabolic problems, and can therefore be used to assess health and welfare status (Dittrich *et al.*, 2019). In the past, on-animal sensors (e.g. 3D accelerometers) have been used for detection of health problems, with varying success. Off-animal sensors such as video cameras have the advantage that a single sensor can be used for: (1) multiple animals simultaneously; and (2) many different behavioral traits, based on spatial as well as temporal features of the individual cows. However, to extract valuable information from the unstructured video footage, targeted data processing is crucial, for example via the use of cow detection, key point identification or pose estimation (Porto *et al.*, 2013).

In this study, we aimed at developing a tool for the detection of 'lying down' (LD) and getting up (GU) events from video footage collected in a freestall barn as leverage for lying behaviour analysis in dairy cows. We started from state-of-the-art cow detection and tracking models, and propose the use of bounding box (bb) properties to predict the LD and GU events.

Materials & methods

Data. Video footage of a freestall barn with synthetic flooring was collected at the research facility 'Dairy Campus' in 2019. Cows were fitted with a 3D accelerometer (IceQube, IceRobotics Inc. Edinburgh, Schotland), from which the timing of GU and LD events was derived. Based on the location of the animals, 1 out of 4 cameras were selected, and the footage surrounding the GU or LD event was isolated to obtain videos of approximately 460 frames (frame rate 15f/s). To collect ground truth information on timing of the GU and LD events, the cows' poses were manually observed and the exact start and end frame were determined and recorded together with the animals' coordinates when lying down. When the target cow was occluded by other cows, videos were removed from the analysis.

Cow detection and tracking. A cow detection model (YOLOv5x) combined with a tracking algorithm (DeepSORT) trained and validated on a separate dataset on the high performance computing infrastructure of WUR (the Netherlands) were applied on each video to obtain bounding box coordinates (x,y), their width (w) and their height (h) for each of the cows (Kamphuis *et al.*, 2022). All further data processing was done in Matlab R2020b (The Mathworks Inc., Natick, MA, USA), and corresponding code can be found on https://git.wur.nl/iadriaens/kb_ddht_ai. Bounding box data of the target cows was gathered using the Euclidian distance of the centre of each bb with the ground truth location of the target cow during lying down (i.e. using either the first (for GU events) or last (for LD events) 50 frames of each clip). When no cow detection had taken place, or when the target cow was not detected in more than 10% of the frames and could not be reidentified, the clip was removed from the analysis. Examples of the bb are shown in Figure 1.

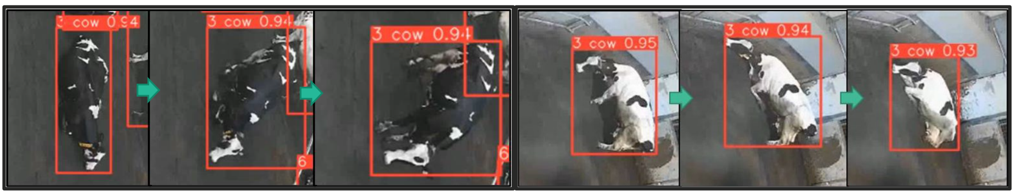

Figure 1. Example of bounding box changes of cow during a getting-up (left panel) and lying-down (right panel) event.

Bounding box properties. The bb coordinates, width and height were smoothed with a third-order Savitsky-Golay filter with window size of 25 frames to correct for noise and instability caused by the YOLOv5 detection algorithm. Next, we calculated the Euclidian distance from the center of the frame (*cntr*), the circumference of the bb (*circ*), and the ratio between width and height (*rwh*) as follows:

$$cntr = \sqrt{(x_{bb} - x_c)^2 + (y_{bb} - y_c)^2} \quad (1)$$

$$circ = 2 * h_{bb} + 2 * w_{bb} \quad (2)$$

$$rwh = \frac{w_{bb}}{h_{bb}} \quad (3)$$

These bb properties (bbp) were standardized with a min-max standardisation as follows:

$$\text{bbp}_{\text{st}} = \frac{bbp - \min(bbp)}{\max(bbp) - \min(bbp)} \quad (4)$$

Finally, the ratios between each of the bbp_{st} (e.g. *centr*/*circ*, *circ*/*rwh*) to represent simultaneous change across bbp were computed.

Event detection and statistics. Unique statistical change points, and local and absolute minima and maxima of the bbp_{st}, of the ratios, and of their first and second derivatives were calculated as proxies for sudden change and maximal slopes in the bbp throughout frames. Next, the location of first and last local peaks within the frame time series was compared with the annotated location of the start and end of the GU/LD events and the frame and time distance was calculated. Appropriate bbp for the detection of GU and LD are those consistently close to the target event throughout all videos, independently of the location and orientation of the cow. The detection was considered sufficiently accurate when the detection was within 2 to 5 seconds (i.e. 30-75 frames) of the actual event.

Results

Eighty-four clips were retained for the analysis, from which 45 with a LD event and 39 with a GU event. The results are summarized in Table 1 with the best performing bbp printed in bolt. When both events were considered, the maximal difference in the *circ* and *cntr* showed the best detection accuracy, with an average distance of resp. 1.3 and 2 seconds from the event, and 82.1% of the videos having a distance of less than 2 seconds from the event. Considering only the GU behaviour, the best detection criteria were the maxima of the second derivatives of *cntr* and *circ*, with average distances of respectively 0.7 and 1.7 seconds and 92.3 to 74.4% of the events being detected within 2 seconds of the ground truth. For LD detection, additional to the maximal differences in *cntr* and *circ*, also the ratios between *cntr* and *circ* and the changepoints of these ratios in all three bbp performed well, with up to 80% of the LD events detected within 2 seconds and 86.7% within 5 seconds of the ground truth.

Discussion

This study investigated the possibility to use a cow detection and tracking algorithm to identify lying down and getting up events via changes in bbp time series over frames. The idea relied on the expectation that during lying down, body shape and location remain constant, while during the GU/LD event, they suddenly change. When the detection algorithm works well and draws its bb tightly and accurately around the animal, the defined bbp were expected to link with lying behaviour. In this study, we showed the potential of using bbp for detection of lying down and getting up events with video footage, and demonstrated that the criteria most suitable for detecting LD or GU separately might differ.

Table 1. Time distance and % events detected within 2 to 5 seconds with the different bounding box properties for all (GU+LD), get-up (GU) and lying-down (LD) events.

	Time distances			% within 2-5 seconds		
	GU+LD	GU	LD	GU+LD	GU	LD
max_dif_cntr[1]	**1.3±3.0**	1.4±3.5	**1.3±2.5**	**82.1-85.7**	84.6-87.2	**80.0-84.4**
max_dif_circ[2]	**2.0±3.7**	2.9±4.6	**1.3±2.5**	**71.4-77.4**	61.5-74.4	**80.0-80.0**
ratiopeak_cntr_circ[3]	4.3±5.2	7.0±6.4	**2.0±2.0**	45.2-65.5	30.8-43.6	**57.8-84.4**
cpts_rat[4]	3.2±4.6	4.6±5.4	**2.1±3.6**	64.3-73.8	48.7-59.0	**77.8-86.7**
der2_cntr[5]	3.4±3.3	**0.7±1.1**	5.8±2.6	46.4-59.5	**92.3-97.4**	6.7-26.7
der2_circ[6]	2.8±3.1	**1.7±2.2**	3.7±3.5	54.8-67.9	**74.4-82.1**	37.8-55.6

[1] maximal difference in centerdistance.

[2] maximal difference in circumference.

[3] peak in ratio between centerdistance and circumference.

[4] statistical change point in ratios across the 3 bbp.

[5] second derivative of center distance.

[6] second derivative of circumference.

In the past, accelerometer sensors have been used to automatically detect lying down or getting up (Ledgerwood *et al.*, 2010), and also in this study, this served as ground truth benchmark. As these sensors are individually attached to the cows, a separate sensor has to be purchased for each animal, which increases costs. Also maintenance (e.g. battery, broken sensors replacements) and the limited number of potential behaviours monitored are drawbacks for this type of technology that are overcome by using cameras. While accelerometers do not require a separate identification step needed for machine vision, these sensors have the disadvantage to not provide spatial information regarding the behaviours monitored.

The advantage of using bbp over e.g. key point based methods, or even over methods that specifically annotate videos for lying animals (Porto *et al.*, 2013) is that it allows a general cow-detection algorithm (for which only bb annotations are needed) to be used as a basis for multiple location and body-shape based behavioural analyses. This limits annotation time and contributes to interpretability and understandability of the methodology, which is an advantage when aiming for e.g. monitoring and decision support. The same bb and tracking could be used for e.g. analysis of feeding behaviour (based on location and orientation of the bb at the feeding rack) or social behaviour (by calculating proximity). The alternative of including the behaviour as input for the (deep learning) video analysis would require many more time consuming and tedious annotation work, with a lower number of potential applications.

Future work can focus on further automatizing the algorithm, testing it on more and longer videos and multiple cows simultaneously. Additionally, a regression model in which several of the bbp time series can be combined potentially further improves detection accuracy.

Video-derived behavioural analysis allows for automated monitoring of multiple animals and multiple traits simultaneously in a cost-effective way using few sensors. In the future, this can contribute to improved monitoring and precision phenotyping, and as such, ultimately leverage a more sustainable dairy sector.

Acknowledgements

This study was part of Knowledge Base Data Driven High Tech (KB-38-008-002), funded by the Dutch Ministry of Agriculture, Nature and Food Quality.

References

Basso B., and Antle, J. (2020) Nature Sustainability 3:254-256. https://doi.org/10.1038/s41893-020-0510-0.

Dittrich I., Gertz M., and Krieter J. (2019) Heliyon 5(11):e02902. https://doi.org/10.1016/j.heliyon.2019.e02902.

Ledgerwood DN., Winckler C., and Tucker CB. (2010) J Dairy Sci 93(11):5129-5139. https://doi.org/10.3168/jds.2009-2945.

Kamphuis C., Adriaens I., Ouweltjes W., and Hulsegge I. (2022) Proc. of the 12th WCGALP, Rotterdam, the Netherlands.

Porto S., Arcidiancono C., Anguzza U., and Cascone G. (2013) Biosyst Eng 115(2):184-194. https://doi.org/10.1016/j.biosystemseng.2013.03.002.

Rushen J., Chapinal N., and de Passillé AM. (2012) Anim Welfare 21:339-350. https://doi.org/10.7120/09627286.21.3.339.