Journal of Systems Architecture 116 (2021) 102139

Contents lists available at ScienceDirect

Journal of Systems Architecture

journal homepage: www.elsevier.com/locate/sysarc

Tracking and analysing social interactions in dairy cattle with real-time locating system and machine learning

Keni Ren [a,c,*], Gun Bernes [b], Mårten Hetta [b], Johannes Karlsson [a]

[a] Department of Applied Physics and Electronics, Umeå University, Håkan Gullessons Väg 20, Teknikhuset, Umeå, Sweden
[b] Department of Agricultural Research for Northern Sweden, Swedish University of Agricultural Sciences, Umeå, Sweden
[c] Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden

ARTICLE INFO

Keywords:
Ultra-wideband
Computer vision
Dairy cows
Social interactions
Machine learning

ABSTRACT

There is a need for reliable and efficient methods for monitoring the activity and social behaviour in cows, in order to optimise management in modern dairy farms. This research presents an embedded system that could track individual cows using Ultra-wideband technology. At the same time, social interactions between individuals around the feeding area were analysed with a computer vision module. Detections of the dairy cows' negative and positive interactions were performed on foreground video stream using a Long-term Recurrent Convolution Networks model. The sensor fusion system was implemented and tested on seven dairy cows during 45 days in an experimental dairy farm. The system performance was evaluated at the feeding area. The real-time locating system based on Ultra-wideband technology reached an accuracy with mean error 0.39 m and standard deviation 0.62 m. The accuracy of detecting the affiliative and agonistic social interactions reached 93.2%. This study demonstrates a potential system for monitoring social interactions between dairy cows.

1. Introduction

Minimising the sources of stress for individual animals is crucial for optimising animal welfare in the modern dairy farm [1]. Competition for limited resources like food and space is a cause for negative stress. In a loose housing system with cubicles, the animals can move freely, but they have to act in competition with the rest of the herd which may imply problems for cows with low social ranking, for example first calvers versus multiparous cows. There are positive interactions between animals within a herd, which reduce stress. Gentle contact, for example, allogrooming (social licking), is one indication of the formation and maintenance of social bonds between individuals [2]. Thus, when studying cow behaviour it is crucial to be able to discriminate between positive (affiliative) and negative (agonistic) interactions. Increased agonistic behaviours can indicate welfare problems, for example when mixing groups, at large group sizes or insufficient space allowance [3]. A higher amount of allogrooming performed by newly introduced individuals may indicate high level of acceptance in the group [4].

To monitor animal conditions and behaviour in large herds, Precision Livestock Farming (PLF) technologies have been developed. They can provide measurements and act as decision support tools to monitor health and thereby help in optimising the animal welfare. Technologies such as automated body condition scoring [5], lameness detection [6] or prediction of calving [7] are good examples focusing on individual welfare measurements. The social dynamics between animals can be studied by using embedded sensor technology, for example, spatial proximity loggers [8] or Ultra-wideband (UWB) technology [9]. However, these sensors can only detect the social behaviour between individuals based on spatial proximity. Previous studies, for example concerning automatic detection of aggressive behaviour in pigs [10], or registration of cows' social interactions in the waiting area before milking [11] propose computer vision methods to detect the social connections between animals. However, these studies focus on detecting activities without connecting them to the identity of each individual. When studying the interactions and competition between individual animals of different groups (for example, higher and lower social rank) in a herd, more types of data, like identification, tracking and social interactions detection need to be taken into the measurement. The task gets difficult to implement due to lack of data interoperability between multiple formats [12,13]. Maintaining the communication between different data platforms and application programming interfaces is also challenging when it comes to data collection in a production environment [14,15]. A PLF system that can perform identification, tracking, and analysing animal behaviour is therefore needed.

* Corresponding author at: Department of Applied Physics and Electronics, Umeå University, Håkan Gullessons Väg 20, Teknikhuset, Umeå, Sweden.
E-mail address: keni.ren@umu.se (K. Ren).

https://doi.org/10.1016/j.sysarc.2021.102139
Received 14 December 2020; Received in revised form 31 March 2021; Accepted 9 April 2021
Available online 15 April 2021
1383-7621/

In this study, we investigate how PLF technology can be used to identify affiliative and agonistic social interactions to minimise potential sources of stress for cows. The objective was to implement a monitoring system to identify, track, and analyse the social interactions between dairy cows in a herd. The system used UWB positioning technology in combination with computer vision technology to highlight the affiliative and agonistic interactions between dairy cows. First, the real-time locating system (RTLS) based on UWB technology tracked individuals in the whole barn. Then, the computer vision system distinguished different social interactions around the feeding area, using Convolutional Neural Network (CNN). The study aimed to develop a robust system to monitor cow social behaviour in crowded scenes with varying illumination conditions, in order to adapt to real-world scenarios.

2. Materials and methods

2.1. Experimental setup and animals

This study has been made possible by the use of the Swedish Infrastructure for Ecosystem Science (SITES), in this case by support of the field station Röbäcksdalen in Umeå, Sweden, which has a dairy herd for experimental purposes. The animals were treated and kept with permission from the Swedish Ethical Committee on Animal Research represented by the Court of Appeal for Northern Norrland in Umeå, Sweden.

The trial was conducted across a period from 20th August to 2nd October, 2018. The research facility under study is located in the north of Sweden, and has a herd of 120 dairy cows (Viking Red), of which around one third are first calvers. About 60 of the cows are kept in an insulated section, consisting of a free stall with cubicles for resting. The floor plan of this section is shown in Fig. 1. It has a rectangular plan of 42 × 13 m with a feeding alley adjacent to a resting area with 62 cubicles arranged in two rows. The stable has 30 Roughage Intake Control^{TM} feed bunks (Insentec B. V. Marknesse, The Netherlands) installed, with automatic recording of feed intake at individual visits. In our study, five of the feed bunks (marked brown in Fig. 1) had limited access only to the ten cows that were chosen for the study. Five of these cows were first calvers. All ten cows could access all five feed bunks. The remaining 50 cows all had access to the remaining 25 feed bunks. The feeding area was open for all cows to access and pass through during the study period. Feed was delivered in the feed bunks seven times a day by an automatic system to ensure *ad libitum* feeding conditions. The feeding system is further described by Hetta et al. [16]. Milking was carried out twice a day in a milking parlour in an adjunct facility.

A real-time locating system based on UWB technology (further described below) was developed and installed to register individual positions continuously. We put collars with positioning sensors on the selected cows under study and tracked them in the whole barn range. Each animal's location, identification, and movement activity (via an accelerometer in each collar) were recorded ten times per minute.

We also deployed an optical system (Axis Q6035-E PTZ network camera) to record the individual feeding behaviour and social interactions around the selected feeding area (Green area in Fig. 1, around 3.5 m × 3 m) including three feed bunks. The camera was installed at a height of five metres above the floor to get a top-view image of the zone, however, the height was not enough to cover all five feed bunks under study. The camera system collected video data of the selected area of the barn continuously for 45 days, with a frame resolution of 1280 × 720 pixels, 25 fps. The videos were calibrated and rectified according to the barn floor plan and synchronised with the RTLS system. The camera view was also used for evaluating the performance of the UWB system.

Fig. 2 shows an overview of the system, with the RTLS module for identification and tracking in the whole barn section and the computer vision module which was developed to detect social behaviour of the cows around the feeding bunks, using Convolutional Neural Network (CNN). These two modules were synchronised to each other with time-stamp and mapping through the floor plan of the barn.

Table 1
The measurement of the anchors.

Anchor	X axis	Y axis	Height above the floor
Anchor 0	22.6 m	1.3 m	2.9 m
Anchor 1	10.6 m	1.3 m	2.9 m
Anchor 2	28.6 m	1.3 m	2.9 m
Anchor 3	34.6 m	1.3 m	2.8 m
Anchor 4	7.9 m	12.2 m	2.2 m
Anchor 5	18.1 m	12.1 m	2.2 m
Anchor 6	28.1 m	12.1 m	2.2 m
Anchor 7	38.4 m	12.1 m	2.2 m

2.2. Real-time location module

We designed an UWB RTLS sensor tag based on the Decawave DWM1000 module. This is an IEEE 802.15.4-2011 UWB compliant module that operates on frequency bands from 3.5 GHz to 6.5 GHz. In order to also record activity of the cow, an ST LIS2DE 3-axis accelerometer was added to the sensor tag. The sensor tags can operate in three different modes depending on their programming:

- Sensor tag that is attached on the collars of the tracked cows.
- Anchor node that is a static reference point in the coordinate system. The distances to all sensor tags in range are calculated from each anchor.
- A special anchor that is called a “zero-anchor”. This device is working as a data provider to the gateway system. The zero anchor has the function for tag synchronisation. There is only one zero-anchor in the positioning system. All tracked sensor tags must therefore be in range of the zero-anchor.

The location calculation includes three steps in the UWB system. First, the tag makes a broadcast that requests every listening anchor to provide information that is used to calculate distances. In the second step, each anchor is sending the requested information within a time slot that is assigned to each of them. The sensor tag is using the provided information to calculate the distance to each anchor. In the third step this information is forwarded to the zero-anchor. The distances between sensor tags and anchors are then forwarded to the gateway machine for calculation of positions. The location calculation is done by Least Squares Estimation (LSE) and both the calculated position and the raw distance data is sent to a backend server and stored in a database.

In this study, we used eight anchors to cover the whole barn section. The plan of the anchor positions is shown in Fig. 1 and the measurements of positions are shown in Table 1.

Although the RTLS system captured the real-time location of cows over the whole barn area, the performance evaluation was selected to be done on ten cows when they appeared in the feeding area within the camera scene. During the test period, three tags went out of function, so eventually our study included seven cows.

To evaluate the RTLS module's performance from the camera, an interface was developed in MATLAB to label ground truth position. The image from the Axis camera was calibrated and rectified according to the barn plan. The camera scene was calibrated by placing a chessboard in the middle of the camera view area at the average cow shoulder height 1.5 m. The lens distortion was removed, and a homography was estimated using Zhang's method [17], which projected each of the camera images onto the cow shoulder plane. The calibrated scene is shown in Fig. 3. The calibrated top-view image was used to evaluate the performance of the RTLS system. The top-view images were mapped using four reference points on each feeding bunk to match

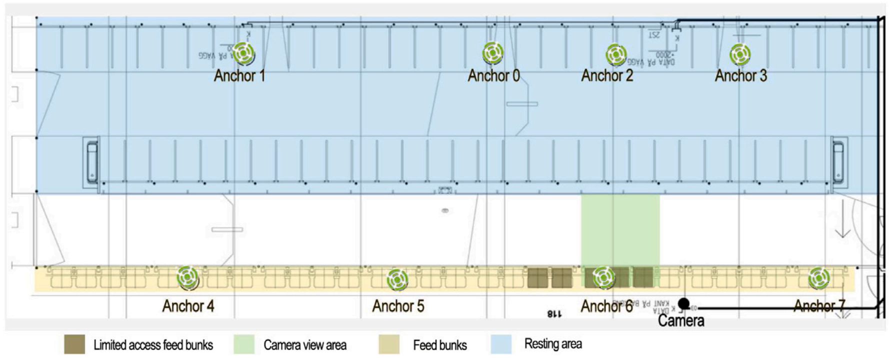

Fig. 1. Floor plan of the barn section under study with the UWB anchor positions. Eight anchors were installed at the positions marked. Five feed bunks, which are marked brown, had limited access for ten cows. The green area was monitored with the camera. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

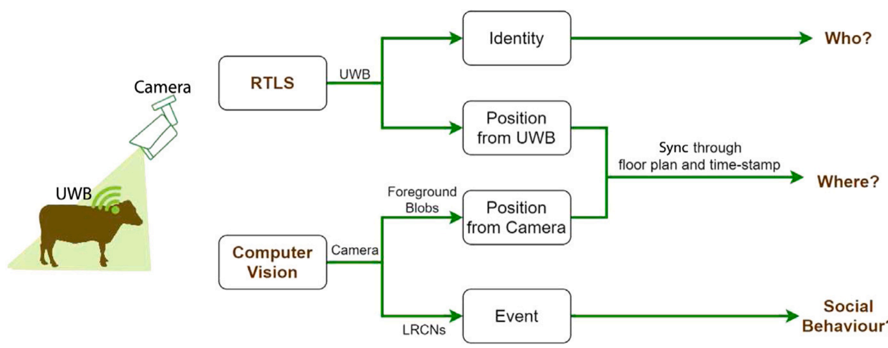

Fig. 2. Overview of the monitoring system.

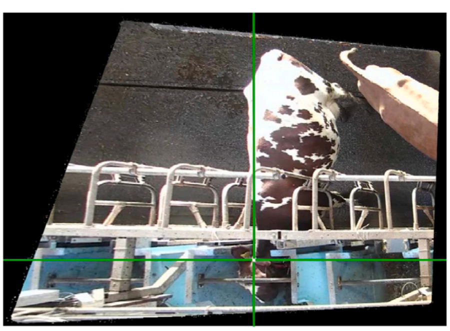

Fig. 3. Example of interface developed to compute the control ground truth localisation of UWB tags on the cows.

the coordinates of the RTLS system. The position of the sensor on the cow's collar was manually identified and marked using the cross of the green lines (Fig. 3) on the top-view image. The marked coordinates were transferred to the barn floor map coordinates. The transferred coordinates were used as the cow's ground truth position.

The real-time location system was evaluated by comparing each tag's position $P_i(x, y)$ with the human labelled position from the top-view images $P_i^C(x, y)$. For each tag, the localisation error ε_i was the distance between the position provided by the RTLS system and the control point verified by the operator:

$$\begin{aligned} \varepsilon_i &= \| P_i(x, y) - P_i^C(x, y) \| \\ &= \| \sqrt{(x_i - x_i^C)^2 + (y_i - y_i^C)^2} \| \end{aligned} \tag{1}$$

Errors from each tag were brought into computing the localisation mean error and standard deviation for the number of the event N where $i = (1, 2, \ldots, N)$.

2.3. Computer vision module

In total 1080 h of videos were recorded in this study, of which videos from 14 days daytime (06:00 to 18:00) with different weather were selected. The amount of video data gave a fair overview of different light conditions, crowded situations, and the diurnal activities. Although the feeding bunks gave limited access for ten cows, the feeding area was open for all 60 cows in the barn section. The videos showed all of the cows appearing in the camera scene.

The videos containing social behaviour event information were segmented and labelled manually. Each event video was between 1 to 30 s. The assessment of dairy cow's social interactions was based on the ethogram adapted from Rousing et al. [18] and Foris et al. [19]. First, all interactions between pairs of cows were labelled as affiliative or agonistic. Then, all the positive interactions were labelled into spatial proximity or gentle contact behaviour. All the agonistic interactions were more specifically labelled into one of four states: threat and withdrawal, body pushing, head butting, and head pressing.

- Affiliative
  - Spatial proximity: Two or more cows stay closer than 3 metres to each other without body contact. Since the monitor area from the camera covers around 3.5-by-3 m area, more than one cow appearing in the image, without body contact, was considered as spatial proximity;
  - Gentle contact: The body of one cow touches the body of another cow, or licking another cow.
- Agonistic
  - Threat and withdrawal: One dominant cow makes slight aggressive movement. The other cow shows avoidance or withdrawal. The cows do not touch each other during the whole process;

- Body pushing: Two cows press body against body;
- Head butting: One cow pushes her forehead (directed blow movement) at the body of another cow;
- Head pressing: Two cows push at each other head to head.

An experienced observer confirmed the presence of social interactions in the video sequences.

The overview structure of the computer vision module is shown in Fig. 4. The process was divided into video input steps, pre-processing, feature extraction step and classification. In the pre-processing step, we performed colour consistency, area of interest masking and foreground detection. Then we harnessed a CNN to extract the features from each frame. Finally, we fed the feature sequences to a bidirectional Long Short-Term Memory (LSTM) architecture to perform the behaviour classification task.

2.3.1. Video pre-processing

The light conditions in the barn varied during the experiment. The indoor illumination was lit most time of the day, but direct sunlight and dark periods of the day made the robust colour-based computer vision task hard to perform. Colour constancy is the ability to show correct colours and eliminate the effect of the colour of the light source. We used the Modified White Patch theory [20] to reduce the effect of illumination variation for its exceptional performance and real-time processing. The Modified White Patch method uses the mean of the highlights by using image pixel sampling instead of the maximal values of the RGB channels of the image. Four 4-by-4 pixels highlight blocks were selected from the scene to calculate the intensity threshold. The corrected pixel values R_c, G_c, B_c, were expressed in terms of original pixel values R,G,B as follows:

$$\begin{bmatrix} K_R & 0 & 0 \\ 0 & K_G & 0 \\ 0 & 0 & K_B \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} R_c \\ G_c \\ B_c \end{bmatrix} \tag{2}$$

where

$$K_R = \frac{WhiteR}{\overline{R_h}} \tag{3}$$

$$K_G = \frac{WhiteG}{\overline{G_h}} \tag{4}$$

$$K_B = \frac{WhiteB}{\overline{B_h}} \tag{5}$$

$\overline{R_h}$, $\overline{G_h}$, $\overline{B_h}$ are the averages of the intensity values of the R, G and B channel over the four selected blocks. WhiteR, WhiteG and WhiteB represent the reference white [255, 255, 255]. The colour consistency was performed directly to the videos. The process was automatic and could be performed as real-time.

In the real production environment, there is usually a complex background appearing in the video stream. The cows under study were Viking Red cows with a red and white coat. They have unique coat patterns, patches and markings with clear contours. Several previous studies have utilised these features to perform visual identification of cows with the use of deep learning [21,22]. The features in coat patterns may affect the detection of social behaviours. In our computer vision module, we wanted to analyse the mechanism of cows' interactions without the disturbance of coat patterns. To solve these challenges, the computer vision system implemented a binary foreground mask for the RGB video streams. The cows as foreground were separated from the background by manipulating the colour components in the CIELAB colour space. The computer vision module detected the red and white components of the cows' coats in the video stream. After testing different light conditions, the colour component thresholds were fixed in the software. The foreground detection performed a binary foreground mask in each frame (Foreground video in Fig. 4). The service alley in front of the feed bunks, where the feeding conveyer and staff could appear, was masked out from the videos. The foreground videos were used for detecting the positions of the cows from the camera view, and their social interactions.

2.3.2. Behaviour detection using long-term recurrent convolution networks

Social interactions between individuals are time series activities. Agonistic behaviours, in particular, which are described as aggressive acts and responses to aggression. For example, an agonistic behaviour started with one cow head butting another cow, then the other cow responded by avoiding or it could result in confrontation and fighting. Instead of CNN, which handles each video frame separately, LSTM was used to better understand the action in the context of the time series information. In our study, video data with the cows' movements over time were extracted. We employed the combination of CNNs and LSTMs, which is referred to as Long-term Recurrent Convolution Networks (LRCNs) [23,24], for the task of social behaviour detection. The CNN layers were responsible for learning image features and the LSTM layers discovered the temporal dependencies. Labelled videos were converted to a sequence of feature vectors using CNN network from each frame. Then the LSTM network was trained on the sequences to classify the video labels. At last, the layers from both networks were assembled to perform the social behaviour classifying task.

Deep learning works well with large amounts of labelled data with modern neural network architectures. In this study, we combined 880 videos sequences and LRCN architectures to test whether deep learning can extract sufficient features to distinguish the differences in cow social behaviour by using computer vision. 460 labelled videos contained affiliative behaviour and 420 contained agonistic behaviour. Of these 880 videos, 623 were used as training partition with 10% as validation partition, and 257 videos were used as testing partition. The affiliative and agonistic interaction videos had equal proportions of training, validation and testing data.

We converted videos to sequences using a pre-trained deep network GoogLeNet (Inception v1) [25] to extract features from each frame feature vectors. GoogLeNet contains multiple inception modules, in which multiple different filter sizes are applied to the same layers to extract features at different scales of detail simultaneously. By learning of diverse types of variations present in the same class of different images, it can reach high accuracy and reduce computational cost [26].

All the videos were resized into 224-by-224 pixels and labelled with one of the defined social behaviours to match the input of the GoogLeNet network. We employed Fine-tuning GoogLeNet to transfer its learned information from the ImageNet [27] domain to cow social behaviour detection task. The pre-trained model was 22 layers deep and contained nine inception modules. Each inception module performed convolution on an input, with different sizes of filters (1×1, 3×3, 5×5). Max-pooling was also performed. The outputs were concatenated and sent to the next inception module. To reduce the model size and the computation, an extra 1×1 convolution before the 3×3 and 5×5 convolutions was added. It used global average pooling at the end of the last inception module. We used the output from the last pooling layer as feature vectors.

The convolutional layers treated each frame of the videos independently. The Bidirectional LSTM (BiLSTM) layer learns long-term dependencies between time steps in a time series video. The BiLSTM layer consists of two LSTMs to take the input from both forward and backward directions. A sequence folding layer was added to the batch of image sequences to assemble both networks. A sequence unfolding layer was added after the convolutional layers to restore the sequence structure of the input. We fed the feature sequence into a LSTM network to perform the classification task. The LSTM network consisted of:

- Feature sequence input layer
- BiLSTM layer with 1000 hidden units
- Dropout layer

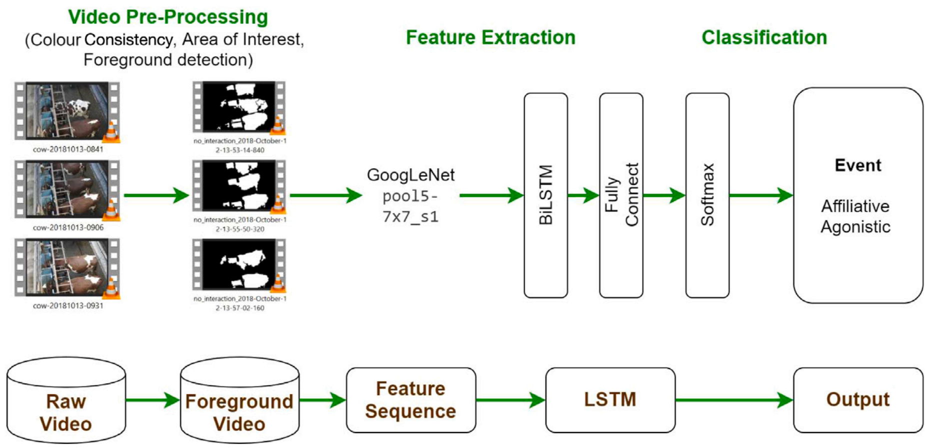

Fig. 4. Overview structure of the behaviour detection.

- Fully connected layer, Softmax layer and Classification layer.

We used the MATLAB software to implement the proposed algorithm, performed on a single CPU laptop. We trained the network with the parameters as mini-batch size of 16, maxEpochs at 20, initial learning rate at 0.0001.

To investigate if the computer vision module can learn more detailed social interactions, the six sub-classes (spatial proximity, gentle contact, head butting, head pressing, body pushing, and threat) of social interactions also were trained using the similar LRCNs model.

To know if LSTM improved the interaction analysing, we also performed the classification using only CNN without LSTM. Transfer learning using GoogLeNet was deployed with sampled foreground video frames as input. We used 18 640 frames as training partition with 10% as validation partition, and 2312 frames were used as testing partition. We transferred the layers to the new classification task by replacing the classification layers according to our data containing six outputs corresponding to the six sub-classes of social interactions. The network was trained with the parameters as mini-batch size of 16, maxEpochs at 20, initial learning rate at 0.0001.

2.3.3. Synchronise RTLS and computer vision modules

Time-stamp was used to synchronise the RTLS and computer vision modules from the timeline. The barn floor plan was used to map the coordinates from both modules. The identities of the cows and their social behaviour were combined by the computer vision system.

After the video pre-processing step, the binary foreground video was automatically labelled by using blob detection, where large groups of connected foreground pixels were considered as a blob. The blob detection helped to register where a cow appeared and the social interaction that happened from the camera scene. The area (number of pixels) and centroid coordinates of each large blob were registered. The blob centroid coordinates were transferred to the barn position according to the floor plan. The individual positions detected by the UWB module around the feeding area under study were assigned to the closest blob coordinates from the computer vision module. Fig. 5 shows an example of an original video frame and the event registered image. The largest blob shows where the body pushing event happened according to the camera scene. The area and centroid coordinated were registered according to the barn map. All UWB tags' positions within the camera scene area were listed in the corner of the image.

3. Result and discussion

The experiment was conducted across a period of 45 days. To give a fair overview of different light conditions and activities during the day, five continuous days were selected to evaluate the RTLS system. Seven individuals with 930 tag positions were evaluated by comparing the position of each tag with labelling results from the calibrated and rectified top-view images. The mean error of all tags' position measurements was 0.39 m and the standard deviation 0.62 m.

Cow identification on farm is frequently achieved by Radio Frequency Identification (RFID) technology [28]. In some research studies, the possibility of using different indoor positioning systems for tracking individuals have been investigated. Wireless Local Area Network technology for cow positioning performed with 1 m as mean error [29] and Bluetooth wireless technology obtained an accuracy of 0.6 m [30]. Porto et al. [31] evaluated the performance of an existing commercially available system, Ubisense system (Ubisense, UK) in a semi-open free-stall cow barn. The localisation mean error of the system was stated to be 0.52 m for the tags applied to the cows. The mean error of our system reached the same accuracy as the commercially available Ubisense system. In this study pure triangulation from all anchors without any filtering was used to calculate the position. Previous studies show improvement in UWB result by adding filters [32] or image analysis [33]. Selecting the four nearest anchors to calculate the distance, instead of all eight anchors, can also improve the accuracy. The four nearest anchors might have a good line of sight compared to anchors far away that might be hidden behind obstacles. Our system gave us easy access to the raw data and the possibility to add other sensors and a computer vision module, compared to if a commercial RTLS system would have been used in the study.

The detection accuracy of the affiliative and agonistic social interactions reached 93.2%. Fig. 6 shows the normalised confusion matrix. Among all the false classifications, 92% of them happened when more than three cows appeared in the scene at the same time, with 15% of the total false classifications happening when more than five cows appeared in the camera scene at the same time. Our study was conducted around the feeding place, as the highest amount of aggressive interactions and allogrooming among loose-housed cattle normally occur in this area [34]. When the feed bunks were newly filled, the feeding area was more crowded, which made the classification of agonistic interactions more difficult.

In our study, agonistic interactions were divided into threat, body pushing, head butting and head pressing, while affiliative interactions were cows spending time next to each other or gentle contact. The average accuracy reached 88.78% in the six classes classification. Fig. 7 shows the normalised confusion matrix for the result of social behaviour detection in the six classes from the full dataset testing. For head butting, body push and spatial proximity, the accuracy reached more than 90%. The head pressing and threat showing behaviours had the lowest detection rates. Both of these events usually last for a very short time (less than 4 s), which may limit the feature learning.

Fig. 5. Example of an original video frame and the event register image.

Actual \ Detected	Agonistic	Affiliative
Agonistic	94.4%	5.6%
Affiliative	7.8%	92.2%

Fig. 6. Normalised confusion matrix of detected agonistic and affiliative interactions.

Actual \ Detected	Spatial proximity	Gentle contact	Head-butting	Head pressing	Body push	Threat
Spatial proximity	90.24%	2.44%	7.32%	0	0	0
Gentle contact	0	87.88%	12.12%	0	0	0
Head butting	1.96%	0	92.16%	1.96%	0	3.92%
Head Pressing	9.09%	0	9.09%	72.73%	0	9.09%
Body push	0	0	0	8.33%	91.67%	0
Threat	6.25%	6.25%	0	0	6.25%	81.25%

Fig. 7. Normalised confusion matrix of social behaviour detection in six classes from full dataset testing.

Earlier studies show positive results of using different types of sensors to analyse social behaviour. As an example, electronic feeding systems can reach the sensitivity 86% for detecting replacement [35]. Guzhva et al. [10] used computer vision to describe social interactions based on geometrical shapes segmented from images. In their study, the result of detecting cows spending time next to each other without body contact had a high accuracy rate (99.9%). But the accuracy of their method of detecting body pushing, head butting and body sniff were 30.5%, 18% and 19.8%, respectively. In a similar way, we also performed the classification into six classes using only CNN on foreground video frame input. The average accuracy was 42.19% with a 97.69% accuracy rate of spatial proximity. The accuracy of head butting and head pressing was 30.0% and 23.5%, respectively, with over 50% false classifications regarding spatial proximity. Our method with LRCNs had a significantly higher accuracy rate, due to adding LSTM to understand the time-series action context.

The proposed system is currently in the development and testing stage. One limitation is the RTLS detection's accuracy, which was not sufficient to give the identities of cows when they were having close body contact with each other. With the computer vision's help, the identities could be recorded as cow IDs at a social interaction event. But the identity could not be assigned to the individual to know which one of the cows that was pushing and which one was the cow being pushed. Moreover, beside the 10 cows with UWB tags, the remaining 50 cows also could appear in the camera scene. Currently, the identity can be used as a reference method for farm management. To improve the synchronisation between RTLS and the computer vision system, RTLS update intervals can be increased to a higher frequency. However, it is a trade-off with battery life. Our future work will concentrate on using the computer vision module to improve the accuracy of the RTLS system. Another limitation is that the camera view only covered a small area of the barn. The application can however be extended to a larger area with a different camera setting.

4. Conclusions

A monitoring system that is capable of tracking individuals and analysing their social interactions has been purposed and tested in this exploratory study. The system integrated RTLS and computer vision modules, maintained the communication between them and combined various formats of data. The RTLS based on UWB technology reached an accuracy with mean error 0.39 m and standard deviation 0.62 m and the detection of the affiliative and agonistic social interactions reached the accuracy 93.2%. The proposed system may help to achieve a real-time automated tool for continuous monitoring of social behaviour in a dairy barn environment. The individual activity and social behaviour in the herd can be used as inputs in an early warning system for the herd manager to detect anomalies in health and welfare of individual cows.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was part of the AnimalSense project. The authors wish to thank the European Union Interreg Botnia-Atlantica who financed this project. We also acknowledge the staff at the field station Röbäcksdalen for their help in the practical part of this project.

References

[1] D. Fraser, Animal welfare: Translating science into practice, in: Advances in Agricultural Animal Welfare, Elsevier, 2018, pp. 129–143.
[2] S. Sato, K. Tarumizu, K. Hatae, The influence of social factors on allogrooming in cows, Appl. Animal Behav. Sci. 38 (3–4) (1993) 235–244.
[3] M.-F. Bouissou, The social behaviour of cattle, Soc. Behav. Farm Animals (2001).
[4] I. De Freslon, J. Peralta, A.C. Strappini, G. Monti, Understanding allogrooming through a dynamic social network approach: an example in a group of dairy cows, Front. Vet. Sci. 7 (2020) 535.
[5] M.F. Hansen, M.L. Smith, L.N. Smith, K.A. Jabbar, D. Forbes, Automated monitoring of dairy cow body condition, mobility and weight using a single 3D video capture device, Comput. Ind. 98 (2018) 14–22.
[6] G.G. Miguel-Pacheco, J. Kaler, J. Remnant, L. Cheyne, C. Abbott, A.P. French, T.P. Pridmore, J.N. Huxley, Behavioural changes in dairy cows with lameness in an automatic milking system, Appl. Animal Behav. Sci. 150 (2014) 1–8.
[7] S. Neethirajan, Recent advances in wearable sensors for animal health management, Sens. Bio-Sens. Res. 12 (2017) 15–29.
[8] N.K. Boyland, D.T. Mlynski, R. James, L.J. Brent, D.P. Croft, The social network structure of a dynamic group of dairy cows: From individual to group level patterns, Appl. Animal Behav. Sci. 174 (2016) 1–10.
[9] L.E. Rocha, O. Terenius, I. Veissier, B. Meunier, P.P. Nielsen, Persistence of sociality in group dynamics of dairy cattle, Appl. Animal Behav. Sci. 223 (2020) 104921.
[10] M. Oczak, S. Viazzi, G. Ismayilova, L.T. Sonoda, N. Roulston, M. Fels, C. Bahr, J. Hartung, M. Guarino, D. Berckmans, et al., Classification of aggressive behaviour in pigs by activity index and multilayer feed forward neural network, Biosyst. Eng. 119 (2014) 89–97.
[11] O. Guzhva, H. Ardö, A. Herlin, M. Nilsson, K. Åström, C. Bergsten, Feasibility study for the implementation of an automatic system for the detection of social interactions in the waiting area of automatic milking stations by using a video surveillance system, Comput. Electron. Agric. 127 (2016) 506–509.
[12] L. Calderoni, A. Magnani, D. Maio, IoT manager: An open-source IoT framework for smart cities, J. Syst. Archit. 98 (2019) 413–423.
[13] C.M. Sosa-Reyna, E. Tello-Leal, D. Lara-Alabazares, Methodology for the model-driven development of service oriented IoT applications, J. Syst. Archit. 90 (2018) 15–22.
[14] J.M. Antle, B. Basso, R.T. Conant, H.C.J. Godfray, J.W. Jones, M. Herrero, R.E. Howitt, B.A. Keating, R. Munoz-Carpena, C. Rosenzweig, et al., Towards a new generation of agricultural system data, models and knowledge products: Design and improvement, Agric. Syst. 155 (2017) 255–268.
[15] R.C. Dobos, D. Taylor, M. Trotter, B. McCorkell, D. Schneider, G. Hinch, Characterising activities of free-ranging merino ewes before, during and after lambing from GNSS data, Small Rumin. Res. 131 (2015) 12–16.
[16] M. Hetta, M. Tahir, C. Swensson, Responses in dairy cows to increased inclusion of wheat in maize and grass silage based diets, Acta Agric. Scand Sect. A 60 (4) (2010) 219–229.
[17] Z. Zhang, A flexible new technique for camera calibration, IEEE Trans. Pattern Anal. Mach. Intell. 22 (11) (2000) 1330–1334.
[18] T. Rousing, F. Wemelsfelder, Qualitative assessment of social behaviour of dairy cows housed in loose housing systems, Appl. Animal Behav. Sci. 101 (1–2) (2006) 40–53.
[19] B. Foris, M. Zebunke, J. Langbein, N. Melzer, Comprehensive analysis of affiliative and agonistic social networks in lactating dairy cattle groups, Appl. Animal Behav. Sci. 210 (2019) 60–67.
[20] E.Y. Lam, Combining gray world and retinex theory for automatic white balance in digital photography, in: Proceedings of the Ninth International Symposium on Consumer Electronics, 2005.(ISCE 2005), IEEE, 2005, pp. 134–139.
[21] W. Andrew, C. Greatwood, T. Burghardt, Visual localisation and individual identification of holstein friesian cattle via deep learning, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017, pp. 2850–2859.
[22] Y. Qiao, D. Su, H. Kong, S. Sukkarieh, S. Lomax, C. Clark, Individual cattle identification using a deep learning based framework, IFAC-PapersOnLine 52 (30) (2019) 318–323.
[23] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, T. Darrell, Long-term recurrent convolutional networks for visual recognition and description, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 2625–2634.
[24] S. Mittal, S. Umesh, A survey on hardware accelerators and optimization techniques for RNNs, J. Syst. Archit. (2020) 101839.
[25] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.
[26] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.
[27] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, Ieee, 2009, pp. 248–255.
[28] S. Ruuska, W. Hämäläinen, A. Sairanen, E. Juutinen, L. Tuomisto, M. Järvinen, J. Mononen, Can stealing cows distort the results of feeding trials? An experiment for quantification and prevention of stealing feed by dairy cows from roughage intake control feeders, Appl. Animal Behav. Sci. 159 (2014) 1–8.
[29] A. Huhtala, K. Suhonen, P. Mäkelä, M. Hakojärvi, J. Ahokas, Evaluation of instrumentation for cow positioning and tracking indoors, Biosyst. Eng. 96 (3) (2007) 399–405.
[30] F.A. Tøgersen, F. Skjøth, L. Munksgaard, S. Højsgaard, Wireless indoor tracking network based on Kalman filters with an application to monitoring dairy cows, Comput. Electron. Agric. 72 (2) (2010) 119–126.
[31] S. Porto, C. Arcidiacono, A. Giummarra, U. Anguzza, G. Cascone, Localisation and identification performances of a real-time location system based on ultra wide band technology for monitoring and tracking dairy cow behaviour in a semi-open free-stall barn, Comput. Electron. Agric. 108 (2014) 221–229.
[32] M. Pastell, L. Frondelius, M. Järvinen, J. Backman, Filtering methods to improve the accuracy of indoor positioning data for dairy cows, Biosyst. Eng. 169 (2018) 22–31.
[33] B. Meunier, P. Pradel, K.H. Sloth, C. Cirié, E. Delval, M.M. Mialon, I. Veissier, Image analysis to refine measurements of dairy cow behaviour from a real-time location system, Biosyst. Eng. 173 (2018) 32–44.
[34] K. Miller, D. Wood-Gush, Some effects of housing on the social behaviour of dairy cows, Animal Sci. 53 (3) (1991) 271–278.
[35] J. Huzzey, D. Weary, B. Tiau, M. von Keyserlingk, Automatic detection of social competition using an electronic feeding system, J. Dairy Sci. 97 (5) (2014) 2953–2958.

Keni Ren is Postgraduate student at the Department of Applied Physics and Electronics, Umeå University and researcher at Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences. Her research area is computer vision and sensor-fusion systems for studying animal behaviour in zoo and livestock production environments.

Gun Bernes is experiment group leader at the Department of Agricultural Research for Northern Sweden at the Swedish University of Agricultural Sciences in Umeå. Since 1989 she involved in different projects concerning feeding and management of small and large ruminants, including behaviour and building issues.

Mårten Hetta (Ph. D) is associate professor in animal nutrition and management at the Swedish University of Agricultural Sciences in Umeå. He has worked with animal nutrition and farming technology with focus on dairy production during 20 years. He is at present appointed as head of the Department of Agricultural Research for Northern Sweden and animal welfare officer for the SLU dairy herd in Umeå.

Johannes Karlsson (Ph. D) is researcher at the department of applied physics and electronics at Umeå university. He has more than 20 years experience from research in wireless sensor networks. He has been project leader for several research projects related to wireless sensor networks for data collection from animals ranging from zoo animals to livestock farming.