Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2018:33
Faculty of Veterinary Medicine and Animal Science

Computer vision algorithms as a modern tool for behavioural analysis in dairy cattle

Oleksiy Guzhva

Computer vision algorithms as a modern tool for behavioural analysis in dairy cattle

Oleksiy Guzhva

Faculty of Veterinary Medicine and Animal Science
Department of Biosystems and Technology
Alnarp

Doctoral thesis
Swedish University of Agricultural Sciences
Alnarp 2018

Acta Universitatis agriculturae Sueciae
2018:33

ISSN 1652-6880
ISBN (print version) 978-91-7760-206-4
ISBN (electronic version) 978-91-7760-207-1

Print: Repro, Alnarp 2018

Computer vision algorithms as a modern tool for behavioural analysis in dairy cattle

Abstract

Looking at modern dairy production, loose housing, i.e. free stalls became one of the most common practices, which, while widely implemented along with different management routines, do not always include the adjustments necessary for assuring animal welfare. The analysis of interactions occurring between cows in dairy barns and their effect on health and performance is of great importance for sustainable, animal-friendly production. The general aim of this thesis was to investigate the possibilities and limitations of computer vision approach for studying dairy cattle behaviour and interactions between animals, as well as take a first step towards the fully automated system for continuous surveillance in modern dairy barns.

In the first study, a seven-point shape-model for describing a cow from the mathematical perspective was proposed and investigated. A pilot study showed that the proposed Behavioural Detector based on the developed shape-model provided a solid basis for behavioural studies in a real-life dairy barn environment.

The second study investigated a classification case from the industry: how animal distribution and claw positioning in specific areas could affect the maximal load on floor elements. The results of the study provided more substantial background data for determining the dimensioning of the strength of the slats.

The third study aimed to take the first step towards an automated system (so-called WatchDog) for behavioural analysis and automatic filtering of the recorded video material. The results showed that the proposed solution is capable of detecting potentially interesting scenes in video-material with the precision of 92,8%.

In the fourth and final study, a state-of-the-art tracking/identification algorithm for multiple objects with near-real-time implementation in crowded scenes with varying illumination was developed and evaluated.

The algorithms forming the multi-modular WatchDog system and developed during this project are the crucial stepping stone towards a fully-automated solution for continuous surveillance of health and welfare-related parameters in dairy cattle. The proposed system could also serve as evaluation/benchmark tool for modern dairy barn assessment.

Keywords: dairy cattle, image analysis, Precision Livestock Farming, computer vision, deep learning, convolutional neural networks, social interactions, tracking, cow traffic

Author's address: Oleksiy Guzhva, SLU, Department of Biosystems and Technology, P.O. Box 103, 230 53 Alnarp, Sweden
E-mail: oleksiy.guzhva@slu.se

Dedication

To my grandpa, who always told me: "There is no shame in not knowing some answers; the shame lies in not knowing where to find them".

“I'll be more enthusiastic about encouraging thinking outside the box when there's evidence of any thinking going on inside it.”

— Terry Pratchett

Contents

List of publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I Guzhva, O., Ardo, H., Herlin, A., Nilsson, M., Astrom, K. and Bergsten, C. (2016). 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. *Computers and Electronics in Agriculture 127*, pp. 506-509.

II Ardo, H., Guzhva, O., Nilsson, M. and Herlin, A. (2016). Cows on concrete slats of the waiting area in a dairy barn estimated by use of image analysis (manuscript).

III Ardo, H., Guzhva, O., Nilsson, M. and Herlin, A. (2017). A CNN-based Cow Interaction Watchdog. *IET Computer Vision 12 (2),* pp. 171-177.

IV Guzhva, O., Ardo, H., Nilsson, M. and Herlin, A. and Tufvesson, L. (2018). CNN-based animal detector with real-time tracking and identification features. *Biosystems Engineering,* (resubmitted after revision).

Papers I, III and IV are reproduced with the permission of the publishers.

The contribution of Oleksiy Guzhva to the papers included in this thesis was as follows:

I Responsible for planning and execution of the work. Responsible for partial analysis of the data and summarising results. The main responsibility for completing the manuscript with input from co-authors.

II Responsible for planning and execution of the work. Responsible for partial analysis of the data and summarising results. Second author responsibility for completing the manuscript with input from co-authors.

III Responsible for planning and execution of the work. Responsible for partial analysis of the data and summarising results. Second author responsibility for completing the manuscript with input from co-authors.

IV Responsible for planning and execution of the work. Responsible for partial analysis of the data and summarising results. The main responsibility for completing the manuscript with input from co-authors.

Abbreviations

AMS	Automatic milking station
CNN	Convolutional neural network
PLF	Precision Livestock Farming
ROI	Region of interest
CV	Computer Vision

1 Introduction

1.1. Preamble

This dissertation consists of several steps or studies aimed towards the one monolithic goal: development of the automatic system for video surveillance and analysis of the behaviour of dairy cows in a modern dairy barn environment.

The main idea of this work was to investigate the possibilities and limitations of computer vision based systems for studies involving dairy cows' behaviour, recognition and tracking of individuals as well as accumulating the knowledge necessary for the practical implementation of computer vision systems in modern dairy production.

The crucial pre-requisite for the success of this work was that the methodology for image analysis (such as object segmentation, recognition, geometrical object model etc.) needed to be developed from "scratch" and be as flexible and robust as possible to function in the real-world situations. This explains the chronological order in which the steps/studies were carried out and also creates the generalised framework for future developments.

1.2. Cows: an introduction

Cows (more often referred to as dairy cattle) are members of subfamily *Bovinae*, most common type of large domesticated ungulates of the genus Bos (*Bos taurus*). Cows have been known and used by humans for different production purposes since the early Neolithic period (10500 BC). The extensive genetic selection and continuous evolution of farm/production environment resulted in some behavioural changes, affecting cows' way of interacting with animal caretakers, each other and farm building environment (McTavish et al., 2013; Decker et al., 2014).

Cows are diurnal animals, and the vision is their dominant sense, they are also able to distinguish between long and short wavelength colours as well as rely more on a perception of moving rather than stationary objects (Adamczyk et al., 2015). One of the other species-typical characteristics influencing the cow's behaviour in different situations is the ability to discriminate between individuals under different circumstances (Coulon et al., 2011). It explains, partly, the complexity of social interactions and hierarchical structure within groups of dairy cows (Kiley-Worthingthon and De La Plain, 1983). There is scientific evidence (Taylor and Davis, 1998) suggesting that cows could use previously stored mental images from earlier social encounters and associate them with real individuals, affecting their intentions and future interactions.

These factors, as well as some other parameters (e.g. breed, parity number, stage of lactation), increase the complexity of planning required for the well-functioning dairy barn and for securing good cow traffic (Barkema et al., 2015).

1.3. Trends in modern dairy production and their effect on animal well-being

A growing population leads to increased demand for milk and meat, influencing the exponential growth and development of a dairy sector. Matters relating to climate change suggest that environmental impacts need to be addressed, as is livestock sustainability (Geers and Madec, 2006). To address these issues, means are required whereby the health and welfare of individual animals could be supported while at the same time maintaining the economies of scale that could be obtained with intensive farming (Halachmi et al., 2000; Hermans et al., 2003). Furthermore, the demand for affordable food means that cost efficiency pressure exists towards the operation of fewer, but more intensive dairy farms that are highly rationalised. These large farms require the tremendous attention to management routines and strategies around animal health and welfare to assure the highest input-output ratio at the lowest cost possible (Olofsson, 1999; Rutten et al., 2013; Miguel-Pacheco et al., 2014). With prices for dairy products varying largely within the EU, the “marginal difference” between economically successful farms and those trying to reach higher income is very small. To maintain the sustainable production without jeopardising the “animal” part of it, farmers required concentrating more on workflow planning as daily amount of their tasks increases (Halachmi et al., 2000; Grandin, 2015). As a result, the animal caretakers (farmers, veterinarians, farm advisors and other actors) have less time to monitor and provide the proper care for each animal (von Keyserlingk et al., 2009).

Concurrent with this development, there is intense pressure from consumers to improve animal welfare, for example by keeping the animals in conditions close to natural ones (Kilgour, 2012; Grandin, 2015). Public health concerns about antimicrobial resistance mean there is an urgent need to reduce the use of antibiotics by improving the health of individual animals.

The well-being of cows depends on many variables related to the barn environment: feeding strategy, cow traffic solutions, number of re-groupings with “new” animals re-establishing the hierarchy, flooring conditions, competition for resources (e.g. water, brushes, concentrate feeders), and the area per cow, cow comfort. As a result, the production system in combination with care strategies determine the environmental variables and therefore, good health and welfare for the cows (Adamczyk et al., 2015; Barkema et al., 2015).

1.4. Farm size and the need for continuous monitoring of animals in specific areas of the dairy barn

With the rapid development of modern dairy production, loose housing, i.e. free stalls swiftly became one of the most common housing alternatives, which, while widely implemented along with different management routines, do not always consider the natural behaviour of dairy cows. The analysis of interactions occurring between cows in dairy barns and their effect on health and performance is of great importance for sustainable production with high standards of animal health and welfare (Phillips, 2002; Grandin, 2015).

According to recent reports (Rutten et al., 2013; Barkema et al., 2015), the average size of the dairy farm is continuously increasing which results in a more substantial number of animals for everyday control and caregiving (per caregiver and farm). As daily farm work includes many different aspects, the time for observing the animals and finding those in need of additional care is dramatically decreased, which could lead to production diseases being unnoticed until later stages, requiring immediate veterinary attention (Geers and Madec, 2006; Barkema et al., 2015). By assuring early detection of diseases and monitoring the health of the animals continuously and in real time, it is possible to increase the end value of the product for the consumer by creating animal-friendly production conditions. Studies are showing (Hermans et al., 2003; Castro et al., 2012) that animals in pain or with the ongoing pathological conditions will express the deviations from their typical behaviours, which could be utilised as a valuable indicator for building the models describing animal's states of well-being. In case of clinical illness, the animal goes through the set of behavioural changes, which could be used by farmer or veterinarians to ease further diagnostics or adjust the management practices accordingly.

For an automatic milking system (AMS) to be efficient, the motivation of cows to enter it becomes one of the most crucial aspects. Feed is the primary motivator to visit the AMS (Prescott et al., 1998; Halachmi et al., 2000). However, the expectation to get feed in the AMS creates competition in the waiting area before the AMS. This competition affects the social interactions, the timing, and the regularity of the visits to the AMS (Melin et al., 2005, 2006). This situation is similar to the competition around an automatic concentrate feeder (Olofsson, 1999; Herlin and Frank, 2007). Thus, monitoring social behaviour and aggression in the waiting area are of interest both from an AMS efficiency perspective as well as for animal welfare.

1.5. Social interactions and their importance for the animal well-being

Among one of the leading challenges for cows in early lactation is their introduction to a new group/part of the barn/system. There are studies (Hedlund and Lovlie, 2015) showing that regrouping is one of the underlying reasons for the occurrence of agonistic behaviours and increased competition for different resources.

Dairy cows are social animals, and the "group" is an essential resource for them and influences health state, welfare, and production performance. In nature, the average size of the herd/group in *Bos*-species is on average 15-20 individuals, which considering the average size of modern dairy herds (50+ individuals at least) increases the complexity of possible social interactions and their effect on cows' well-being. Studies (Kikusui, Winslow and Mori, 2006) show that a secure and positive social environment in cow barn results in both short- and long-term effects linked to different aspects of cow health, as well as production performance, longevity, and lowered stress levels on group/herd level (Rault, 2012).

The maintenance of the social environment in modern dairy herds is exceptionally complex activity and not only is time demanding but also requiring a sound understanding of underlying reasons and motivational factors influencing cows' behaviour (Vieira, de Pasille and Weary, 2012; Shin, Kang and Seo, 2017).

There is the growing interest within the research community towards the "social-buffering" in farm animal welfare and health (Boyland et al., 2016). This concept is underlying the importance of dyadic relationships and hierarchical structure within a group of dairy cows and all the interindividual relationships as well as their effects on the functionality of farm environment. The defying principles and analysis of these relationships include not only direct interactions occurring between the individuals but also take into consideration such parameters as proximity between animals, animal distribution and their spatial preferences (Gygax et al., 2010).

There is evidence (Seyfarth and Cheney, 2015) showing that the consequences of regroupings and changes in social groups of dairy cows and their linkage to possible health and stress-related deviations are not always considered in everyday farm practice. Therefore, these problems require more solid knowledge to create the set of strategies for assuring environment oriented towards the animal welfare (Kondo et al., 1989).

The occurrence of agonistic interactions within the established groups of dairy cows is directly linked to some new individuals entering these groups, as studies suggest (MacKay et al., 2014; Boyland et al., 2016). Prolonged conflicts and need to re-establish previously formed social hierarchy/dominance order could escalate and lead to severe injuries and even affect the safety of the working environment in the dairy barn (Proudfoot, Weary and von Keyserlingk, 2012). To understand the underlying reasons for this, one should remember that cows form preferential social bonds (Bergman and Beehner, 2015) and these bonds could vary depending on the resource or situation, as well as actual health state of the individual. There are different indicators (social interactions expressed/performed) that could indicate the degree of involvement of the individuals and thus serve as important assessment parameter for animal caretakers or veterinarians. With increasing knowledge and scientific evidence related to the importance of social environment for dairy production, there is the demand for information on optimal management strategies (e.g. group size, stocking density, the optimal number of re-groupings and such).

1.6. Flooring as one of the housing factors influencing cows' behaviour

One of the substantial factors related to animal well-being is the knowledge of how the housing and management in a dairy barn could affect each other and animals (Barkema et al., 1998; Schreiner and Ruegg, 2003). It is essential for farmers to be aware of how the building and construction of the dairy barn will influence cows' behaviour and health in relation to production performance.

Among one of the most crucial housing parameters having the direct relation to the ease and will of cows to perform/express behaviours is flooring (Telezhenko and Bergsten, 2005). The floors in modern cow barns often consist of concrete, which is harsh and abrasive. Cows spend a substantial amount of their time walking or standing on different surfaces, making it easier to qualitatively and quantitatively assess the parameters having the direct impact on health or performance (e.g. lameness, mounting behaviour, comfort) (Telezhenko, 2007).

Studies have shown that increased social and agonistic activity in modern livestock housing systems exacerbates the leg problems and lameness (Cook, Bennett and Nordlund, 2004). Lameness in cows is common and a serious concern, especially for first calving heifers. Currently, too many young heifers are culled before even reaching the reproductive state. Many of these animals are also treated with antibiotics with questionable results and when there are non-antibiotic alternatives available (Bergsten et al., 2017). Such a development does not correspond with values such as resource efficiency and sustainability, and it is questionable whether the use of painkillers may be more efficient. Thus, there is a need to improve flooring in group housing facilities for cows, to reduce the risk of leg problems by prevention strategies rather than merely treating the problem with antibiotics (Vokey et al., 2001).

Slatted concrete floors are commonly used in dairy barns for aisles, feeding and waiting areas, e.g. before milking. The design of slatted floors should include good drainage capacity obtained by the slot width or the void ratio (Magnusson et al., 2008; Platz et al., 2008). While ensuring the good drainage capacity of the floor by adjusting void ratio, the adequate claw support should also be kept in mind. One of the direct effects of changing the void ratio will be increased or decreased hygiene, which could be then correlated with several health/welfare-related parameters (e.g. risk of mastitis, the risk of leg injury due to slipperiness, cleanliness of animals and so forth). The construction of the slats must consider this together with the length of the slats and the load from the weight of the animals on the slats to dimension the load strength. Presently, the calculation of the strength of the slats is set by a European standard which is entirely adopted by Sweden (SIS, 2007). The loads used in the calculations are based on the type and mass of animals and put into load classes. Three variable characteristics loads shall be taken: vertical characteristic linear and point loads and a horizontal characteristic point load. However, the calculation of load strength considers a twin or a multiple slat construction instead of the prevalent single beams used in Sweden. This is the situation where the application of computer vision techniques for tracking cows in the specific areas of interest could provide valuable information on spatial distribution of individuals and help in the evaluation of construction elements (e.g. actual load on concrete beam).

1.7. Precision Livestock Farming (PLF) and computer vision as modern solutions for continuous animal surveillance

According to Whates (2007), the Precision Livestock Farming (PLF) approach is "The application of the principles and techniques of process engineering to livestock farming to monitor, model and manage animal production."

The housing of cows in large groups requires a sound knowledge of the cow's basic social behaviour, an ability to monitor and understand the needs of individual animals within the group and appropriate care interventions to prevent health problems (Dominiak and Kristensen, 2017). However, so far no practically useful technology to monitor individual animals in new, sometimes "de-synchronised" environments, is available (Giot, El-Abed and Rosenberger, 2013; Redmon et al., 2016). Therefore, a sensor-based PLF approach could become an efficient aid for the animal care providers and provide tools for automatic surveillance of the animals and pre-clinical indicators of health problems (Rutten et al., 2013). Monitoring of individual animals through sensor technologies within different production systems and care strategies, in combination with health recordings, may accumulate crucial data about essential principles for good animal health and animal well-being. This knowledge could also help in reducing the stress levels among cows (Tullo et al., 2016).

Due to the rapid development of the dairy industry and overall complexity of everyday farm work (e.g., advanced feeding and management techniques, heat detection, health control on individual and heard levels, etc.), visual assessment of individuals by farmer becomes ineffective and requires large investments of time (Busse et al., 2015). Interestingly, the first attempts at investigating the possibilities of computer vision approach for livestock production were made already in the late 80s, showing the potential for further development (Marchant, 1988a). However, certain levels of hardware development, as well as the established methodology for visual signal processing, were needed to establish the full potential that image technology could bring to animal surveillance (Van der Stuyft et al., 1991).

There are a number of ways to use image analysis in the animal production, e.g.: automatic monitoring of locomotion (Miguel-Pacheco et al., 2014), oestrus detection (Tsai and Huang, 2014), interactions within sexually active groups (Sveberg et al., 2013), pig aggression (Oczak et al., 2013) and animal density in a poultry house (Kashiha et al., 2013). However, problems with identifying and differentiating individuals on mixed backgrounds with nonhomogeneous illumination for the existing segmentation methods slow the development of feasible solutions for commercial farming. Therefore, new techniques for object tracking and recognition are needed.

1.8. Available solutions for animal tracking and identification

The use of different computer vision based systems for animal tracking and monitoring in dairy barns is rapidly developing an area within PLF research field. There are solutions suited for segmenting animals from the background in different areas of a dairy barn and distinguishing between different behavioural states (e.g. standing, laying down, moving) (Porto et al., 2015). Combining these methods with advanced machine learning approach (Simonyan and Zisserman, 2015), it is possible to create an algorithm capable of complex scene evaluation and multifactorial analysis of a dairy barn environment in relation to the desired hypothesis.

To monitor farm animal's behaviour and assess all the occurring interactions, one should be able to quantify and qualify performed interactions in a reliable, repeatable, and continuous manner (Cangar et al., 2008; Porto et al., 2013; Guzhva et al., 2016). The focal observations and manual analysis of the recorded video material are two of the most common approaches used for these purposes. Such manual approach is time-demanding and is largely based on a skill of the person performing the annotations and interpretation of the performed behaviours (Haidet et al., 2009). Another important feature is the ability to correctly identify the animals in overly crowded scenes, under varying illumination, during different hours of the day.

The need for robust identification of individuals has become a multi-dimensional problem involving monitoring of production performance as well as individual health and the well-being of animals in dairy herds (Dziuk, 2003; Tullo et al., 2016; Busse et al., 2015; Carne et al., 2009). During the past decade, several alternatives for animal tracking and identification were proposed: WI-FI-, RFID-, GPS-, ultra-wideband and Bluetooth-based products (Ahrendt et al., 2011; Nadimi et al., 2012; Rutten et al., 2013; Awad, 2016). Among all methods mentioned above, RFID-modules gained significant popularity over the course of past years due to certain advantages over the other methods. These advantages include the enormous potential for data storage, affordability and scalability, relatively long battery life. However, nevertheless all the advantages, RFID-modules do still require a considerable amount of work for setting them up: manual marking of animals with RFID-tags, protocols and infrastructure, integration into existing on-site digital ecosystem (Busse et al., 2015; Carne et al., 2009). Therefore, considering the increasing average size of dairy herds and number of individuals requiring monitoring, there is a need for a flexible and non-invasive system capable of alternative ways for individual tracking and identification (Banhazi and Tscharke, 2016).

As one of the alternatives, computer vision systems could ensure more frequent sampling, larger sequences recorded and analysed (Cangar et al., 2008; Sellers and Hirasaki, 2014; Tullo et al., 2016). One of the other benefits of using computer vision system is the flexibility of the recording setup and many features that could be extracted from the video material and used for descriptive analysis of the behaviours, locations of animals, identification and more (Guzhva et al., 2016). In a case of real-time monitoring and analysis, the need for extensive storage capacity is also resolved, as video stream could be assessed directly, making the procedure more efficient and suitable for practical on-farm use.

1.9. Importance of dialogue in cross-disciplinary projects and possible pitfalls

For any animal-oriented project that involves technology and necessitates cross-disciplinary approach, the first and the most important issue is communication. Where animal scientists see behavioural issues, animal-environment related problems and different physiological aspects defining the animal as a dynamic system, engineers or mathematicians see equations, data structures and hardware related problems. Communication barriers and lack of “agreed-upon” terminology are the main reasons for slow progress in such projects often oriented towards the state-of-the-art development and complex support solutions.

As the personal reflection after four years in a project involving different aspects of animal science as well as mathematical and computer sciences sprinkled with recent advances in computer vision and deep learning, one should not underestimate the importance of simple dialogue. Being lucky enough and having insights and practical experience from both camps, there is a need to admit that the definition of "animal well-being" becomes almost unbearable task while trying to consider all the modelling aspects. Animals are dynamic individual systems, with many parameters influencing their behaviour, therefore making the classical "black box" approach less viable when it comes to continuous analysis of animal health and welfare. To obtain that sound knowledge required for even simple forecasting and decision support systems, one should include a number of factors that often come from different disciplines.

Seeing colleagues with "classical" university education covering one, sometimes two relatively narrow fields, suffering from an inability to explain their hypothesis to somebody working on a different side of the academic umbrella, underlines the importance of interdisciplinary communication. The progress and advances in such field as Precision Livestock Farming require clear understanding and definition of familiar concepts (e.g. animal, behaviour, well-being, sensor, algorithm, or model).

2 Aim of the thesis

The general aim of the project was to investigate the possibilities and limitations of image analysis approach for studying dairy cattle behaviour and movement as well as take a first step towards the fully automated system for continuous surveillance in modern dairy barns.

The aim of paper I was to investigate social interactions between cows in the waiting area with four AMS units and to collect data by three top-down view video cameras.

The paper II aimed to estimate the presence of claws on individual slats by observation of animal distribution on the slatted floor in a waiting area to robotic milking. The study was performed to give more substantial background data for determining the dimensioning of the strength of the slats.

The paper III aimed to take the first step towards an automated system for behavioural analysis and filtering of the recorded video material.

The paper IV aimed to create flexible, state-of-the-art tracking/ identification algorithm for multiple objects with near-real-time implementation in crowded scenes with varying illumination and based on recent progress in computer vision and convolutional neural networks.

3 Materials and methodology

3.1. Animals, housing, and study area

All studies (Papers I-IV) were carried out at the commercial free-stall dairy barn situated in Smedstorp, Skåne County, Sweden.

There were 250+ Swedish Holstein cows during the time of studies that had access to the Region of Interest (ROI). The cows were milked in four automatic milking stations (AMS), (VMS®, DeLaval, Sweden) with a pre-selection of cows for milking using a milk-first traffic system. The practical implication of milk-first system is that animals need (depending on time since the last milking session) to go through the AMS unit if they want to access feeding area of the barn.

The common waiting area (6 x 18 meters) before entering any of the four AMS was used for all the studies. With average (according to the statistics from VMS) milking rate of 2.4 per animal per day, the rough estimate for daily passage rate through the study area was 600 cows. The cows had free access to all four AMS at any time during the day. The floor in the waiting area consisted of the concrete slats with 125 mm in width and the slat opening of 35 mm.

3.2. Development of video acquisition method and recording setup

Video recordings were made using three Axis M3006-V (Axis Communications®) cameras with a broad view angle of 134 degrees. They were placed in the ceiling of the barn at the height of 3.6-meters, pointing straight down to optimise overview of the study area. There was a significant overlap between the camera images to avoid missing events taking place at the border between the cameras. See Figure 1 and 2 for some example frames. In total 24 months (only 4 + 2 months out of those 24 were used for analysis) of video-material was recorded with the frame resolution of 800x600 pixels, 8-bit RGB colour space and a framerate of 16 frames per second to achieve a balance between the size of the video file and the quality of the image. The videos were recorded using the H264 codec for compression.

Figure 1. Example frames from the recorded video material.

Figure 2. The frames from Figure 1 projected onto the cow shoulder plan and stitched together to form the overview over the entire waiting area.

3.3. Calibration of cameras and plane estimation of the study area

To fulfil the specific research questions regarding dairy barn slatted floor and distribution of animals on it, the correct transition of image coordinates into real-world coordinates was crucial. To assure that all the coordinates were reliable and that the observer/algorithm was able to identify all the objects correctly, three images were merged and synchronised after applying normalisation algorithms. Setups with only one camera (even with wide observation angle) could suffer from some image artefacts (e.g. radial distortion, tangential distortion, occlusion between objects/cows) therefore; it was decided to use three cameras for a relatively small ROI in the waiting area.

The classical pinhole camera model augmented with a lens distortion model was used to model the camera. The camera setup was calibrated by placing markers on the walls and stands in the middle of the waiting area (see Figure 3).

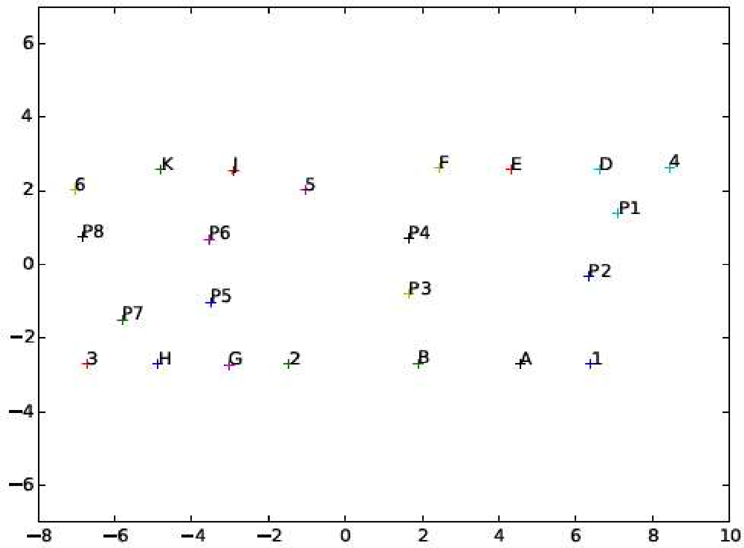

Figure 3. The location of the calibration markers on the virtual floor (1.49 meters above the real floor) used to calibrate the cameras. The crosses indicated marker positions and next to the cross is a label used to identify a specific point.

They were all placed at the same height and thus defined a plane. The mean cow height in the barn was measured, and the plane was placed at cows' shoulder height. This height was estimated to be 1.49 meters with a standard deviation of 0.05 by measuring twelve random cows in the study area. This was the plane in which all of the landmarks considered below, except for the head, were expected to be found. By projecting detected landmarks back and forth between the camera images and this plane, detections from different cameras could be matched. In addition to the markers, the focal length and lens distortion parameters were provided by the camera manufacturer. The lens distortion was removed, and a homography was estimated that projected each of the camera images onto the cow shoulder plane. At the borders between the cameras, the image became strange as cows there were viewed from different directions on opposite sides of the border. However, this image was only used for illustration. There was enough overlap between the images to allow them to be processed one by one and then the resulting detections could be combined using this calibration. By using homogeneous coordinates, the pinhole camera model that forms 2D image pixels, $x = (x_1, x_2, 1)$, by projecting world 3D points, $X = (X_1, X_2, X_3, 1)$, using a camera matrix, P, could be formed as:

$$\lambda x = PX = K(R\,t)X = \begin{pmatrix} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{pmatrix} (R\,t)X \qquad (1)$$

where f is the focal length of the camera and (p_x, p_y) its principal points, while t and R define its 3D position and orientation. This projective image was then distorted using some lens distortion function:

$$\hat{x} = f_{dist}(K^{-1}x) \qquad (2)$$

Here a radial fisheye lens distortion model was used. By using nominal image coordinates, $\lambda x_n = K^{-1}x$ and $\hat{x} = (\hat{x}_1 - p_x, \hat{x}_2 - p_y, 1)$, the inverse of this distortion function could be expressed as:

$$|x_n| = \tan(\sum_i k_i |\hat{x}_n|^i) \qquad (3)$$

where k_i is the set of distortion parameters and $|x| = \sqrt{x_1^2 + x_2^2}$. Note that with this distortion model, the focal length becomes part of the distortion and Equation 3 allows the pixels produced by the camera, x, to be transformed into nominal projective coordinates, x_n, directly. The distortion parameters, k_i, was provided by the manufacturer and the principal point, (p_x, p_y), was assumed to be at the centre of the image.

In the camera images, the markers placed in the cow shoulder plane were allocated manually and their distorted coordinates, $\hat{x}$, were registered by clicking on them. Also, real-world distances between the marks were estimated using a laser distance meter. From these measurements the world coordinates, x_s, were calculated using multidimensional scaling (Young and Householder, 1938). One homograph, H, for each camera was fitted to the point correspondences allowing the pixels to be projected onto this plane using:

$$\lambda x_d = Hx_n \qquad (4)$$

This generated a common coordinate system for all of the cameras which allowed detections from each of the cameras to be projected into this common frame. That way different cameras could be used in different parts of the waiting area. To minimise the amount of occlusion taking place, each part of the waiting area should use the closest camera to get a view from as straight above as possible. For that, the camera position in the cow shoulder plane needs to be estimated. That could be achieved by RQ factorisation of H (Hartley and Zisserman, 2004),

$$H = K_d R_d = \lambda \begin{pmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} R_d \tag{5}$$

were the camera position is given by c = (c_x, c_y, 1). Figure 2 shows an image where each pixel has been chosen from the camera closest to that pixel, i.e. with minimum $|x_d - c|$. The final crops used from each camera consisted of the pixels in this stitched view, with a border of 75 pixels added on each side to make the overlap 150 pixels, which roughly corresponds to the length of one cow.

3.4. Data preparation and development of seven-point shape model for behavioural studies (Paper I) and WatchDog/Tracker algorithms (Paper III and IV)

To assure that data in video sequences with behaviours was representative for the whole period of recordings and to assure the variability of the behavioural events, we used a script randomising the choice of frames in the pool with video sequences. This script randomly picked up sequences of frames with potentially interesting events (total number of annotated frames in all the selected sequences, n = 8370, with an interval of five frames between every annotated frame) from the whole set of 288 h of videos. An experienced observer confirmed the presence of social interactions in the video sequences. To prepare images from the sequences for further analysis, manual frame segmentation was used. To increase the speed of the initial segmentation, an interactive segmentation tool for Matlab proposed by Gulshan et al. (2010) was used. This tool used a new state-of-the-art approach based on the improved geodesic star convexity method. The social interactions were identified based on the alignment of geometrical shapes segmented from the image and positively identified as cows by experienced observers (Figure 4).

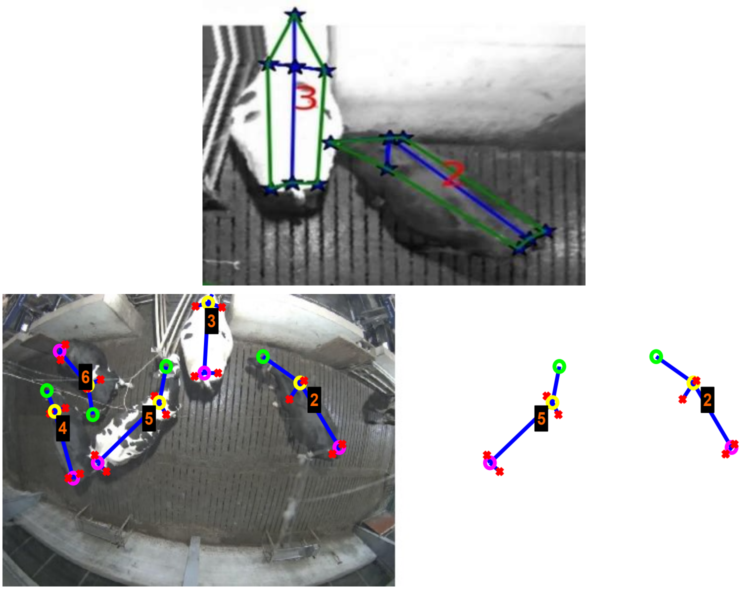

Figure 4. Example of the shape model (landmark points assigned to every animal).

To achieve accurate identification of behaviours by both observer and algorithm, seven landmark points (head, left and right shoulder, front middle, left and right hip and back middle) were then manually assigned to every cow present in an image. The information from every landmark point, containing image coordinates and absence or presence of occlusion, was stored in .JSON files, named to match the unique image code with camera number, date and time stamps. One group of the segmented object pixels with the segmentation index value was defined as a cow. The identification of social interactions was based on the ethogram adapted from Rousing and Wemelsfelder (2006) and all the interactions between pairs of cows were annotated into one of five states:

- Body sniffing: one cow was having her muzzle close (stretched towards) or touching the body of another cow;
- Body pushing: one cow was pressing (or staying very close) her body against the body of another cow;
- Head butting: one cow was having her forehead positioned (directed blow movement) towards the head of another cow;
- Head pressing: one cow was having her forehead pressed to the head, neck or body of another cow;
- No interaction;

The CNN was trained using stochastic gradient descent with momentum (Rumelhart et al., 1986). An initial learning rate of 1.0 was used, and it was lowered by a factor 1/10 each time the validation loss flattened out. A batch size of 256 and a momentum of 0.9 was used. The network was regularised using weight decay of 0.0001 and batch normalisation (Ioffe and Szegedy, 2015). The CNN-detector was trained from the manually annotated data. Utilising the full set of annotated frames (n = 8370) to get unbiased results, training and testing of the detector was performed by using 10-fold cross-validation. A classification system, investigating features extracted from pairs of cow's shape models, was developed (Figure 4). Every cow shape model went through Bookstein normalisation (Bookstein, 1997) and every pair of cows was investigated independently, via symmetric training, forming two training examples for input. From every training example, positions and distances for every landmark point were extracted (24 positions plus 49 distances, resulting in 73 features per training example) and used for investigating possibilities of interactions. Furthermore, an outline of each shape model formed by six line segments and the closest point to each point on the other cow was found. A single number for each such closest point was formed by unfolding the outline into a straight line and rescaling each segment to place the connecting endpoints of the segments at the integer positions 0, 1, . . . 5.

In each frame, each ordered pair of cows was the source of one potential interaction. Our interaction-detector investigated each pair of cows, and the output of this detector was a probability, that indicated the likelihood of the detector for this particular pair of cows performing each interaction. When the detector was certain that there was an interaction going on, this detection probability was close to 1. If the detector was certain that there was no interaction, it was close to 0, and when it was uncertain, it was close to 0.5. In a scene with six cows, there would be 30 ordered pairs investigated. This gave 30 different probabilities:

$$p_i \text{ for } i = 1 \ldots 30 \tag{6}$$

These were combined into a single probability (one for each behaviour) for the entire frame giving the probability that any of the pairs was in a state of, e.g., 'Head Pressing, which gave the detection probability for the frame:

$$p = 1 - \prod_i (1 - p_i) \tag{7}$$

3.5. CNN cow detector architecture

The detector was split into two steps. The first step (the landmark CNN) is a fully convolutional CNN that detects the landmarks in the image. The second step (the cow CNN) is another CNN that uses the probability map produced by the first CNN as input and detects the cows and their orientation (Figure 5).

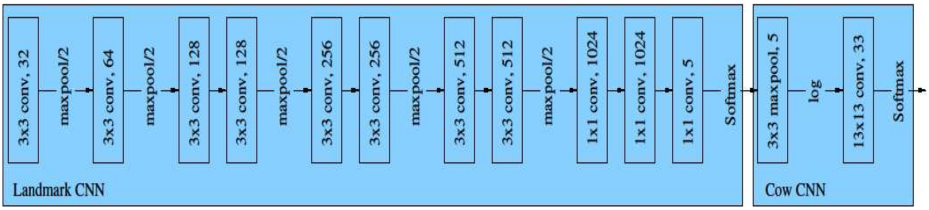

Figure 5. The architecture of the cow detector.

Only four of the landmarks were used for the landmark CNN: head, front middle, cow centre and back middle. The discarded landmarks left and right shoulder are close to the front middle while the discarded landmarks left and right hip are close to back middle. This was done because these landmarks do not provide much extra information about the cow's position or orientation and including them would slow down the experiments.

Also, due to the max-pooling used in the CNN, the output feature map has a too low resolution to separate such close landmarks well, which means that they might compete for the same pixels in the output probability map. That might harm performance.

However, extending to use more landmarks is straightforward. The architecture of this network is a fully convolutional CNN similar to VGG (Simonyan and Zisserman, 2014) with batch normalisation (Ioffe and Szegedy, 2015) after each convolution step. That means that instead of fully connected layers at the end, 1×1 convolutions were used. Only valid outputs from the convolutional and maxpool layers were kept. Applying the network to a high-resolution image would apply the landmark detector to every position in the image, resulting in an efficient implementation of a sliding window detector. It could be applied to images of resolution $118 + 32n_w$ times $118 + 32n_h$ and would produce an output image of resolution $n_w \times n_h$ with five channels. Each of the channels contains the detection probability of each of the five classes (four landmarks and one background class).

A pixel at (x, y) in the output probability map corresponds to landmark detection at position $(32x + 75, 32y + 75)$ in the input image. During training, $n_w = n_h = 1$ was used, and the net was trained on patches of 150×150 pixels extracted from the input images. The positive examples were centred on the landmarks and randomly jittered ±16 pixels (as the distance between output pixels is 32 input pixels). Negative patches were selected at centres more than 32 pixels from any landmark. In addition to the positive and negative patches a set of "do not care" patches was selected at random centres at distances between 16 and 32 pixels from landmarks. The ground truth (Golden Standard manually confirmed by an observer) probability of these patches belong to the class of the landmark was set to 0.5, and the probability that they are ground was set to 0.5. In some cases, several landmarks appear within 32 pixels of the patch centre. In that case, the probability mass was distributed uniformly among all involved classes. Also, all patches were randomly rotated ±180 degrees. The weights of the convolutions were initiated using random samples drawn from a Gaussian distribution truncated at 2σ, with standard deviation $\sigma = \sqrt{\frac{2}{n}}$, where n is the number of inputs (He et al., 2015). Once the net was trained, the step length of the last maxpool layer was reduced from two to one to increase the output resolution. After that pixels at (x, y) in the probability maps correspond to detections at (16x + 75, 16y + 75) in the input image. During testing n_w and n_h depend on the size of the input frame which varies from camera to camera.

The second step is another fully convolutional CNN that works with the probability map produced by the first CNN as input and tries to detect the cows and their orientations. The full circle is divided into 32 equally spaced orientations which generate 32 different oriented cow classes. In addition to that, there is the "no cow" class, which makes the total number of classes of this CNN 33. The input probabilities were turned into log likelihoods as it made more sense when summing them together. Then the network consisted of a single 13 × 13 convolutional layer.

The same resolution was kept after each layer which means the relationship between pixels coordinates in the output map and the original input frame was the same as for the landmark CNN. The net was then applied to the fully rectified training images producing probability maps of 44 × 46 × 5 pixels. These were used as training examples for the cow detection net (without splitting them into patches). Random rotations ±180 degrees and translations ± 16 pixels were applied to both input images and their annotations. This means that all the 6399 annotated cows could eventually be used as a positive example to each of the 32 orientation classes. The output ground truth probability maps of 44 × 46 × 33 pixels were constructed from the annotations by projecting each cow, i, centre point into the probability map as (x_i, y_i) and calculate its angle a_i as the angle of the line between front middle and back middle landmarks. Then a binary 44 × 46 × 33 mask B (x, y, c) is formed, containing a background mask:

$$B(x, y, 32) = \begin{cases} 0 \quad \text{if} & \begin{matrix} \lfloor x_i \rfloor \leq x \leq \lceil x_i \rceil \\ \lfloor y_i \rfloor \leq y \leq \lceil y_i \rceil \end{matrix} \\ 1 \quad \text{otherwise} & \end{cases} \tag{8}$$

and 32 orientation masks:

$$B(x,y,c) = \begin{cases} 1 \quad \text{if} & \begin{array}{c} \lfloor x_i \rfloor - 1 \le x \le \lceil x_i \rceil + 1 \\ \lfloor y_i \rfloor - 1 \le y \le \lceil y_i \rceil + 1 \\ \text{adist}\left(\frac{2c\pi}{32}, c_i\right) < \frac{2\pi}{32} \end{array} \\ 0 \quad \text{otherwise} \end{cases} \tag{9}$$

for $0 \le c \le 31$ and all i. The adist function calculates the absolute angular distance between two angles. The ground truth probability masks are then produced by normalising B to sum to 1 for each pixel. Finally, the network is trained using the same hyperparameters as for the landmark CNN.

3.6. Classification of claws on slatted floor models (Paper II)

After the network produced the detections, overlapping detections were removed by a pruning state. The Figure 6 below shows some results from the detector where the detected cows are marked with rectangles and the detections removed by the pruning are marked in red.

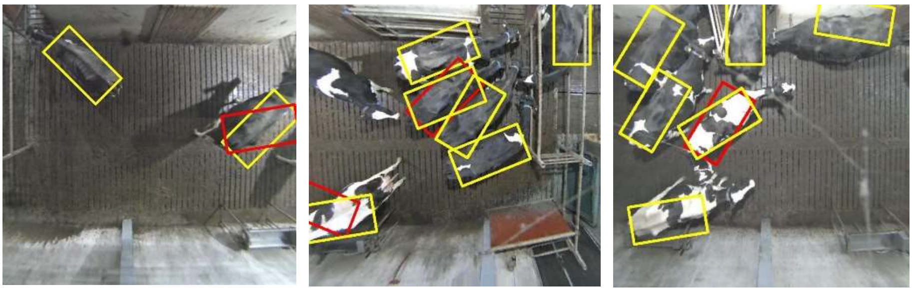

Figure 6. Example of the output image from the cow-detector.

The cow detector only generates a centre position and orientation of each cow. To estimate the position of the claws from such detections, a statistical model was formed from the manual annotations. All the annotated cows were normalised by translating their centre to (0,0) and rotating them to align their body with the x-axis. The normalised positions for all four claws were then plotted in different colours in Figure 7 below, and a mean cow shape was estimated by taking the mean position of each claw. This mean shape was plotted as a red H-like shape with endpoints of the lines marking the claw positions.

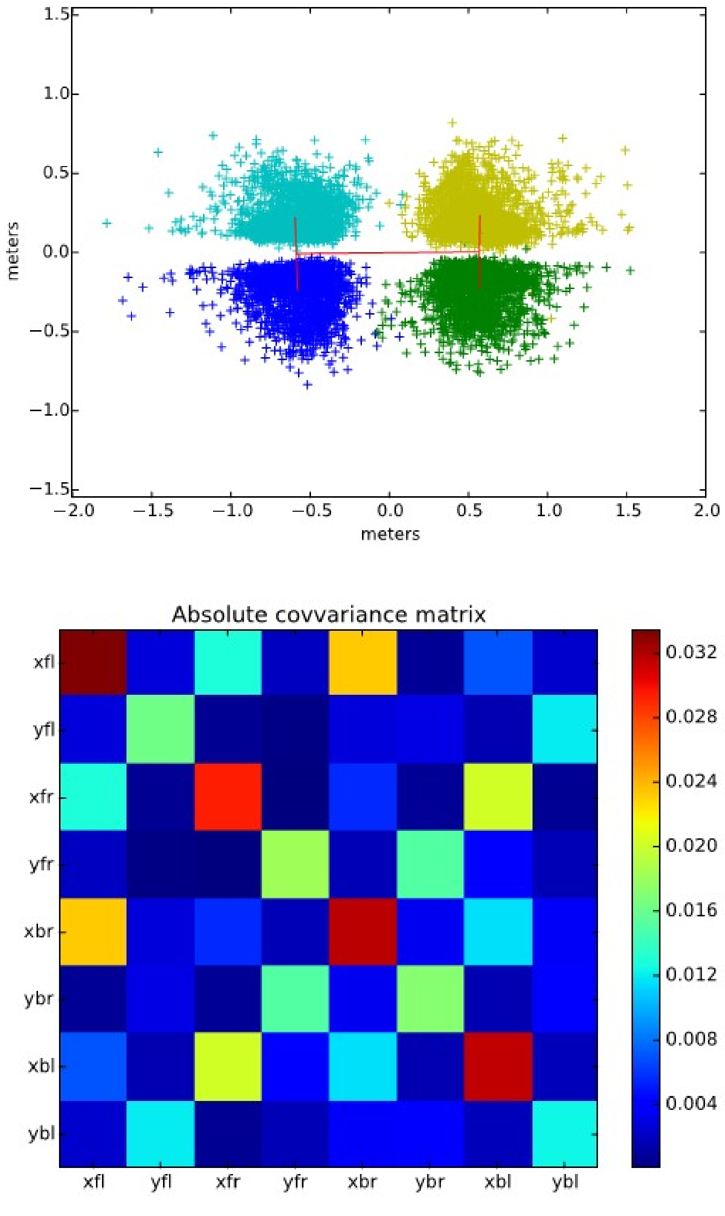

Figure 7. Normalisation positions for all four claws and their covariance matrix.

Figure 7 shows that the distribution of the claws given the cow position is heavily skewed, that means that fitting, for example, a Gaussian distribution to it, will not give a perfect fit as the Gaussian is symmetric.

Also, the positions of the claws are highly correlated to each other as could be seen in the covariance matrix whose elementwise absolute values are plotted above. It shows that the correlation between the coordinates of the claws, (xfl, yfl) and (xfr, yfr) for the front claws as well as (xbl, ybl) and (xbr, ybr) for the back claws are when compared to the variances on the diagonal. Instead of trying to find a parametric distribution that could be fitted well to this data, a non-parametric approach was used. The entire set of annotated and normalised cows was stored as a representation of the claw distribution. To sample from this distribution, a random cow was drawn from this set.

To simulate a floor with 125 mm wide slats and slot openings of 35 mm, a grid with 160 mm wide rectangles was placed in the image (Figure 8), representing the area of the load of a claw on individual beam.

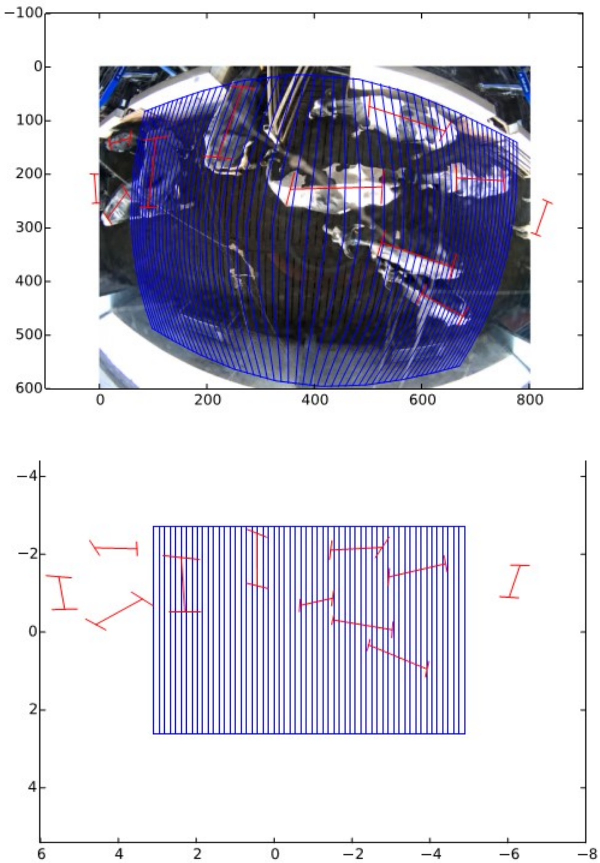

Figure 8. Projection of the virtual floor over the barn floor. The floor is represented by a blue grid with one grid element for each 160 mm slat+slot. The annotated cows are shown as red H-shapes where the endpoints indicate the positions of the claws. It is also possible to project the annotated cows from the images into the coordinate system of the virtual floor, as is shown in the second row of the figure.

The rationality here is that if a claw was placed over the opening, the full weight supported by that claw would still be placed on the slat. Different heights and placements of the rectangles were investigated. The number of claws placed in each rectangle was then calculated and divided by the total number of observed images. This gave a probability distribution over the number of claws on a random slat at a random point in time. This analysis was performed both using the manually annotated cows (1722 images) and using automated detections on a separate set of images (5861 images) that was not used during the training of the detector. For each of the cows detected by the CNN detector, a random sample was drawn from the claw distribution presented above. This sample was then translated and rotated to be placed at the detected centre at the detected orientation. By comparing those results, the precision of the automated process could be estimated. From the probability of there being exactly n claws on a random slat at a random time, p(X=n), the probability of there being n or fewer claws can be estimated as a sum from 0 to n:

$$p(X \leq n) = \sum_{i=0}^{n} p(X = i) \qquad (10)$$

The probability distribution of the worst case at any random point in time is found by taking the maximum over all the slats. On a floor with m slats, the probability of there being n or fewer claws on the maximum is that same as there being n or fewer claws on all the slots. If the slots are assumed to be independent this can be estimated as:

$$p(\max(X) \leq n) = p(X \leq n)^{m} \qquad (11)$$

3.7. Proposed tracking algorithm (Tracker)

The Tracker optimises over sequences of detection likelihoods produced by the CNN and is thus able to utilise all the information provided by the CNN, using per frame non-maximum suppression. The tracking algorithm used the probability map produced by the CNN directly, without first constraining it to a few discrete detections. The probability map consists of probability, $d_{s,t}$, of a cow being detected in each of discrete sets of possible states, s $\in$ $\mathbb{S}$, in frame t.

These states typically consist of the location of the objects (i.e. the coordinates of the probability map produced by the CNN), but could also be more informative as in the case above where the detector also detects the orientation of the cows. Each state, s $\in$ $\mathbb{S}$ then consists of a position (x; y) and an orientation α, i.e. s = (x; y; α) for some discrete sets of | $\mathbb{S}$ | possible states.

The proposed tracking algorithm does not depend on the structure of those states and below $\mathbb{S}$ refers to a general discrete set of states. The only assumption made about the states is that two different objects could not be in the same state at the same time, which makes sense as the position of the object typically is the part of its state.

The state space was augmented with a probabilistic motion model that described how the state of an object was allowed to move from one frame to another. This model was defined as a probability distribution, $p(s_t|s_{t-1})$, over states s_t in frame t given the state of the object, s_{t-1}, in frame t-1. Any such model could be used, but typically the model would assign high probabilities for the object to retain its current state or move to a neighbouring state, while it assigns low probabilities to it jumping further away.

The gates described above were used to indicate when objects enter or leave the scene. Each gate was associated with a specific state. When an entrance gate, with state s_{in}, indicated that a new cow has entered the scene, a new object was instantiated with state s_{in}. Also, when an exit gate with state s_{out} indicated that a cow has left the scene, the object that currently is most likely to be in state s_{out} was removed from the scene. This means that the remaining parts of the tracker could operate under the assumption that the number of objects stayed known and constant from one frame to the next. For each state $s \in \mathbb{S}$ the tracking algorithm could maintain $o_{s,t}$, which is the identity of the object that is currently most likely to have the state s and $p_{s,t}$, which is the probability that the object $o_{s,t}$ has state s in frame t. These values were updated recursively by assuming that $o_{s,t-1}$ and $p_{s,t-1}$ are known and for each state s calculate the most likely previous state:

$$e_s = \arg\max_{\hat{s}} p_{\hat{s},t-1} p(s|\hat{s}) \tag{12}$$

This allows $o_{s,t-1}$ to be propagated using:

$$o_{s,t} = o_{e_s,t-1} \tag{13}$$

To propagate the probabilities, the observation probabilities, $d_{s,t}$, produced by the CNN detector are used:

$$\tilde{p}_{s,t} = d_{s,t} p_{e_s,t-1} p(s|e_s) \tag{14}$$

These propagated probabilities will no longer sum to one. By assuming that the object is still present and its state is one of the states for which it is currently the most likely object, a probability distribution for the current frame could be formed by normalising the propagated probabilities:

$$p_{s,t} = \frac{\hat{p}_{s,t}}{\sum_{\hat{s}|o_{s,t}} \tilde{p}_{\hat{s},t}} \tag{15}$$

The second part of that assumption is an approximation. For distant objects it is insignificant, but for close objects, it might affect the results. Finally, the current state of each object, o, is estimated as:

$$s_o = \underset{\hat{s}|o_{s,t}=o}{\arg\max}\, p_{\hat{s},t} \tag{16}$$

3.8. Real-ID from passive data-markers

As for the identification of individuals, tracking algorithm presented in this study utilises data-markers already integrated into modern robotic dairy barn environment. All the manufacturers producing equipment for automatic milking systems have RFID-tags on animals for accessing selection gates, milking stations, feeders. This means that the information capable of identifying the individual cow is already present and saved in the computer logs every time animal moves/takes action. By combining these passive data-markers with a robust visual tracking system, non-invasive identification of individuals in different situations/parts of the barn made possible. As the ow of animals is usually controlled by the system of selection gates and there are at least several entry points to the scene of the interest, the opportunity to back-trace the real-id number is usually higher with a larger number of registrations per animal. The gates register when cows enter or exit the scene, and this information is together with the identification of the cow is passed to the tracker, which tracks the cow's movements while in the waiting area. However, for this study and to further investigate possible limitations of the proposed system, only one registration at automatic milking station was used for identification. The tracker detected and followed cows to the entrance to the milking station, where the system read the real-ID number. The detector then assigned this real-ID number to a detected cow and followed her along the tracklet backwards to the moment of actual entry to the waiting area.

4 Results

The results section aims to present findings from different modules that form the WatchDog system and follow the logical as well as the chronological order of system development. These modules were developed and tested in the following order: an algorithm for the behavioural analysis (Behavioural Detector module), an algorithm for virtual/real floor assessment and claw placement (Floor/Claw module), an algorithm for filtering video material (actual WatchDog module) and at last an algorithm for tracking and identification (Tracker" module).

4.1. Behavioural Detector module performance

The results showing the accuracy of the detector for two sets of behavioural events (Head Pressing and Body Pushing) within the tested sequences of frames are presented in Figures 9 and 10. The overall performance of the detector and accuracy for all five classes of behavioural events are presented in the form of confusion matrix (Table 1).

Table 1. "Behavioural Detector" module performance while distinguishing between five different behavioural states (no interaction – no_int, body pushing – body_push, head-butting – head_butt, head pressing – head_press, body sniffing – body_sniff) and compared to Ground Truth values.

CLASSIFICATION RESULTS

GROUND TRUTH		no_int	body_push	head_butt	head_press	body_sniff
	no_int	**6719**	1	0	0	2
	body_push	471	**217**	5	16	3
	head_butt	225	0	**52**	11	0
	head_press	318	18	11	**96**	3
	body_sniff	107	0	0	55	**40**

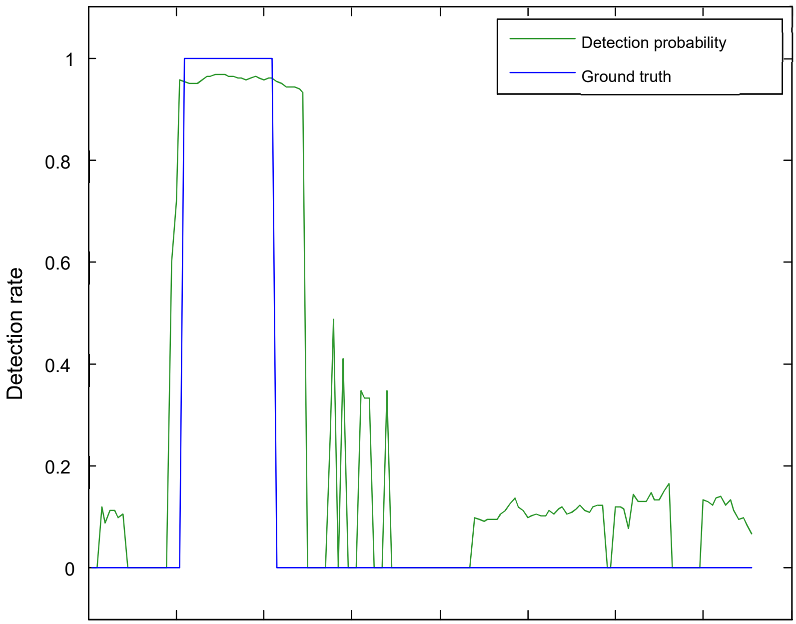

Figure 9. The accuracy of the detector for the Head Pressing behaviour in a test fragment.

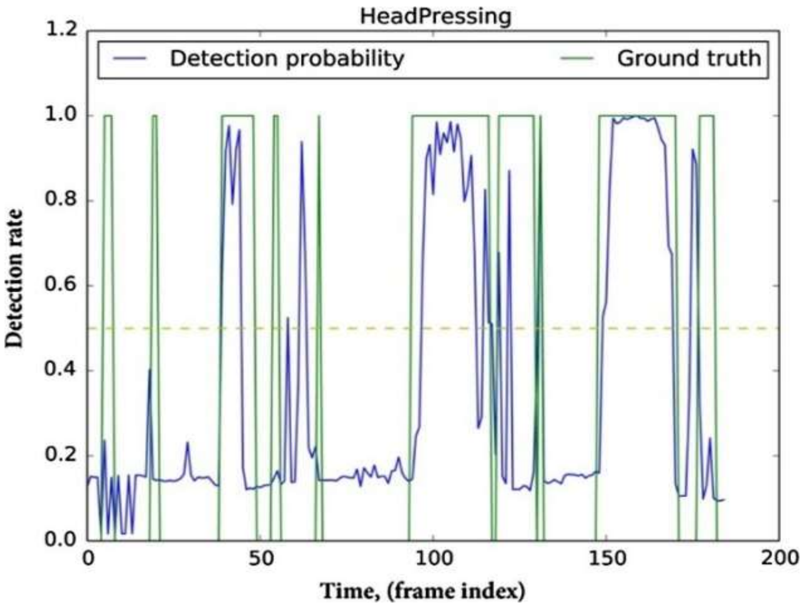

Figure 10. The accuracy of the detector for the Body Pushing behaviour in a test fragment.

4.2. Floor/Claw module performance for slatted floors with different properties (case study)

The first study compared 160 mm wide slats with 560 mm wide slats. The area studied was 16 meters wide, which means that 100 slats were needed to cover the entire area if they were 160 mm wide, while 29 was enough for the 560 mm case. The slats were assumed to cover the entire height of the area (5.3 meters). Experiments were performed both with manually annotated cows and with automatically detected cows. The manual cases consisted of 1722 images randomly sampled from a random camera at a random time. This resulted in a total of 14930 hooves and 77847 slat observations for the 160 mm case and 23079 observations for the 560 mm case. The automated detections were performed on 5861 images resulting in 508960 hooves and 2631490 (23079) slat observations for the 160 (560) mm case. Results are shown in Figures 11 and 12 below. Results are presented both for 160 mm slats and for 560 mm slats.

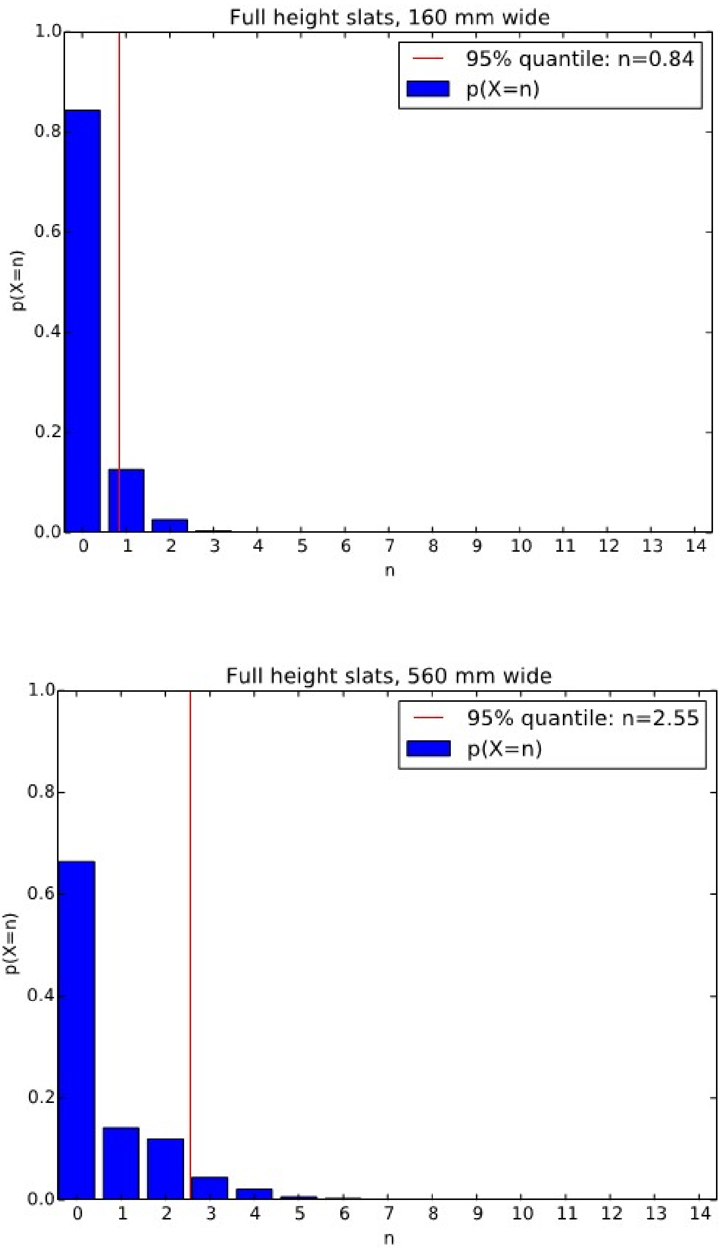

Figure 11. The claw distribution per slat per unit of time modelled for 160 mm (top) and 560 mm (bottom) scenarios.

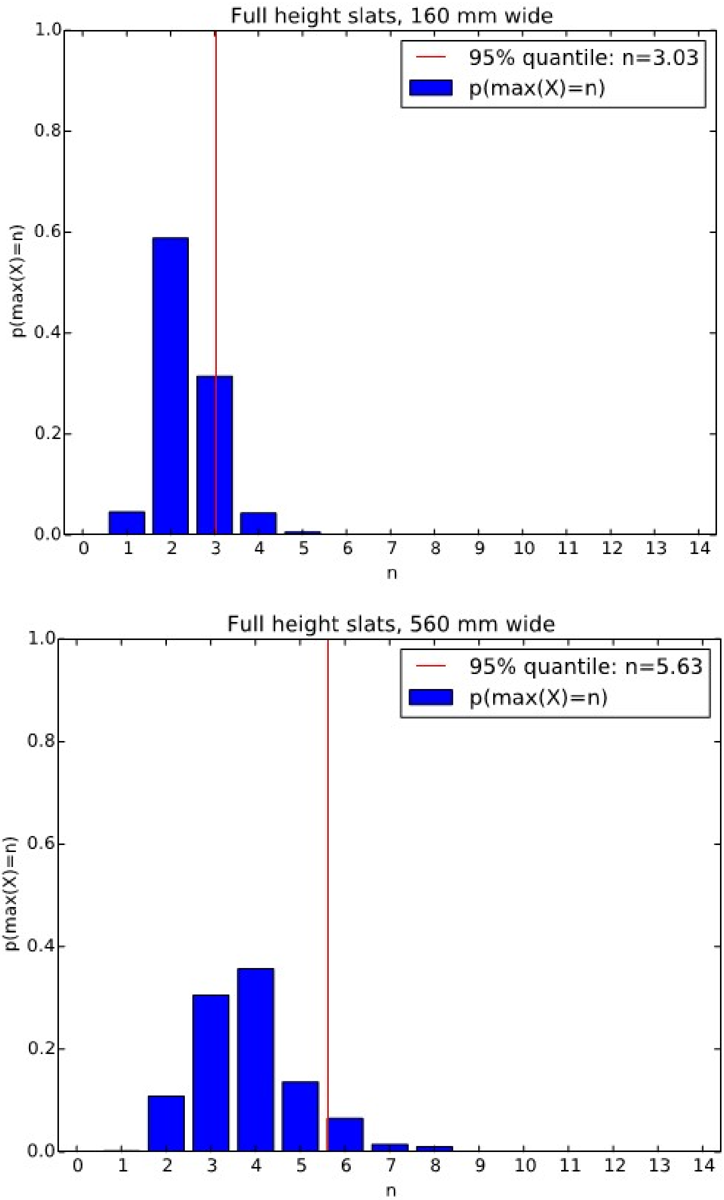

Figure 12. The maximal load on a slat with n or less claws modelled for 160 mm (top) and 560 mm (bottom) scenarios.

The automated results correspond very well with the manual versions, and the narrower 160 mm slat-zones received significantly fewer claw-placements as would be expected. The automated approach slightly underestimates the probabilities as compared to the manual approach. This is probably because the detector sometimes could fail to detect a cow.

4.3. WatchDog module evaluation results

4.3.1. Cow detection

To evaluate the system performance, 6400 frames spread over the entire test-recording were processed by the CNN. It was not ensured that none of the training frames were chosen here, but since they both were chosen randomly from a set of 400 million frames, the chance that they are mutually exclusive is 97.2%. A simple algorithm extracting frames containing two or more cows was implemented. That would be the most basic requirement for interaction, and already this simple criterion, discarded 38% of the recordings. To verify that the discarded video was uninteresting, 500 random frames selected by the watchdog and 500 random frames discarded by the watchdog were automatically annotated using the CNN results and studied manually.

Cows intersecting the borders were ignored in the sense that the images were considered correct regardless of whether such border cases was detected or not. A detection was considered correct if its rotated bounding box overlapped more than 50% of the cow back. A histogram of the number of cows detected per frame is presented in Figure 13.

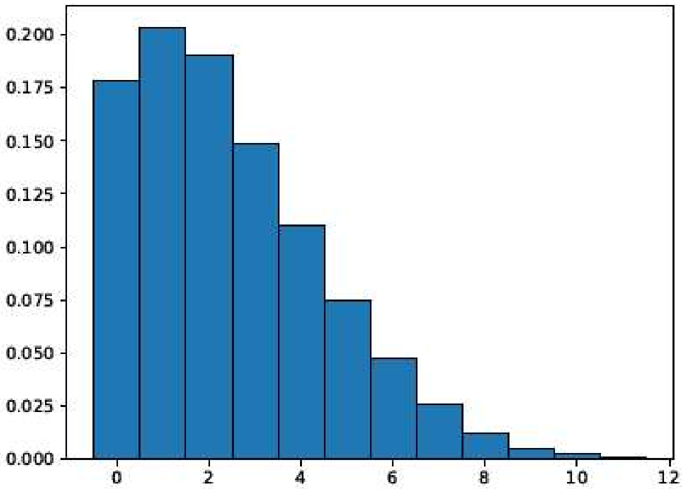

Figure 13. A histogram of the percentage of cows detected per frame. Frames with 0 or 1 cows were considered uninteresting, which according to this detector is 38% of the frames.

Some example detections are shown in Figure 14. Two of the reasons for mistakes are inter-cow occlusion and the combination of landmarks from different individuals.

The evaluation runs at 6.55 fps on a single Tesla K20m GPU, using single precision floats.

Figure 14. Top row: Three images correctly interpreted (all cows detected and no extra detections). Bottom row: The three images were the errors were made (one missed cow and two extra detections).

4.3.2. Interaction detection

To remove even more of the uninteresting video, an additional feature, the minimum distance between the cows in the scene, was also extracted. Then a short sequence containing a lot of interesting interactions consisting of 187 frames uniformly sampled over 10 minutes was extracted and manually annotated by an expert. Five different interactions were manually identified: body pushing, butting, head-butting, head pressing and body-sniffing. Frames where any of these interactions were present, were considered interesting and all other frames uninteresting. The minimum cow distance, d_f, for each frame f, was extracted. It is believed to be a useful feature as cows need to be close to interact. Also, the cows need to be in close proximity for some period of time, so we take the maximum over nine consecutive frames and use as a feature, $x_f = \max_{f-4 \leq I < f+4} d_i$ for the frame f.

By thresholding x_f a simple detector is formed. By choosing different thresholds, the results could be varied between not detecting any uninteresting frames and not missing any interesting frames. The per frame ROC curve in Figure 15 shows the amount of the interesting frames detected (true positives) as a function of the amount of the uninteresting frames remaining (false positives) for different thresholds. It is, for example, possible to discard 20% of the uninteresting frames while only losing 3% of the interesting once. If it is acceptable to lose more of the interesting frames, even more of uninteresting frames could be discarded as detailed by the curve.

An interaction will consist of multiple frames, and in many situations, it is enough to detect a single one of those frames as the user then could be allowed to also look at adjacent frames. Figure 15 also contains a per interaction roc curve that shows the amount of the interactions for which at least one frame was detected (true positives) as a function of the amount of the uninteresting frames remaining (false positives). It shows that it is for example possible to discard 35% of the uninteresting frames without losing any of the interactions.

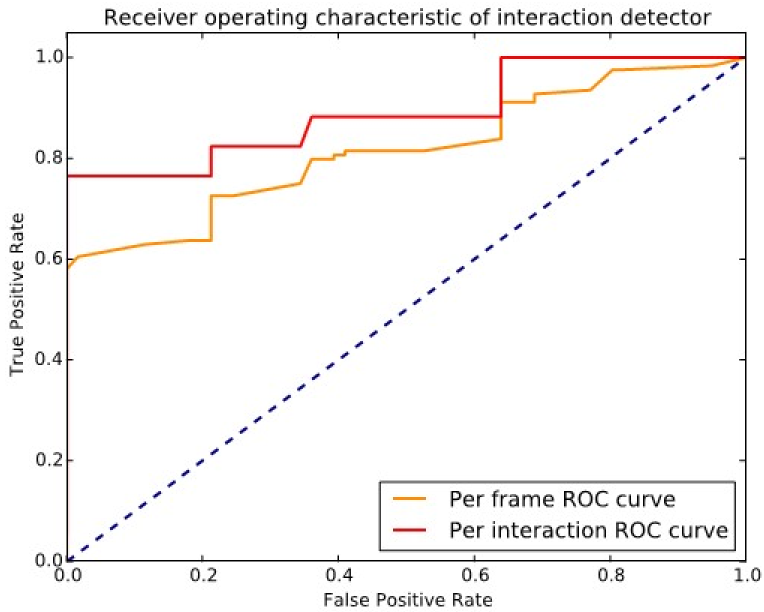

Figure 15. Results for different thresholds, with the number of interesting frames (in orange) or interactions (in red) kept (true positive rate), plotted as a function of the number of uninteresting frames kept (false positive rate).

Note that this is in addition to the filtering by the number of cows present. By combining the two filters, 50% of the uninteresting frames could be discarded while only losing 4% of the interesting frames. If one frame per interactions is enough, it would be possible to discard 60% of the video without losing any interactions.

The system was evaluated on several levels: cows, frames and interactions. Single cows were detected with a hit rate of 97% and a false alarm rate of 2.9%. Single frames were perfectly interpreted 92.8% of the time. Frames with only zero or one cow present were correctly identified 99.5% of the time. All interactions were detected while discarding 60% of the video as uninteresting.

4.4. Tracker module performance

To evaluate the tracking system, two one hour recordings were chosen. One recording with only a few cows in the waiting area during the night (with artificial lighting only) and another recording from a crowded scene (during the day when the sun shines in through the window, Figure 16). The exit time and gate found by the tracker for each cow, that both entered and exited the scene during the recording, were compared with the exit times produced by the selection gates.

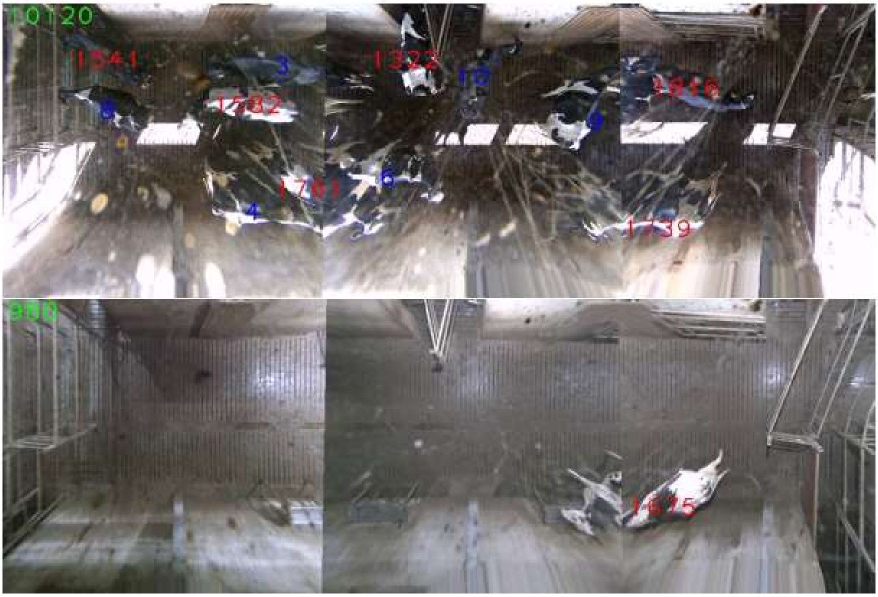

Figure 16. Example frames, with tracked objects marked, from the crowded, sunny (top) and easy (bottom) sequence. The red ID-numbers are initiated and assigned by the selection gates and placed on the correct cow by the tracker, while the blue numbers are placed manually in the first frame and then further tracked.

This difference could be up to 60 seconds even for the correct tracks, as one of the gates was located outside the visible area. Results are shown in Table 2 and Table 3 respectively.

Table 2. Complete trajectories of the simple sequence with columns indicating: cow id-number, tracker found the correct exit gate, time-difference between tracker exit and exit registered by the gate in seconds and the total length of the track in seconds.

COW-ID	CORRECT EXIT	GATE DIFF. (SEC)	TRACK LENGTH (SEC)
1832	1	2.38	26.06
1662	1	8.12	46.88
1733	1	6.44	137.44
328	1	3.88	170.81
1553	1	4.06	374.94
631	1	5.19	86.44
1761	1	42.88	73.94
1562	1	2.50	227.00
1852	1	56.12	129.19
1758	1	2.62	37.50
1803	1	22.94	27.06
1833	1	12.38	71.81

Table 3. Complete trajectories of the crowded sequence with columns indicating: cow id-number, tracker found the correct exit gate, time-difference between tracker exit and exit registered by the gate in seconds and the total length of the track in seconds. The rows below the line are cases when the tracker fails. Within parenthesis is the length of the track in seconds that was successfully tracked before the tracker failed.

COW-ID	CORRECT EXIT	GATE DIFF. (SEC)	TRACK LENGTH (SEC)
1582	1	1.62	1240.44
1739	1	25.38	1212.25
1390	1	5.19	360.31
1549	1	3.12	248.94
1767	1	0.75	173.94
1612	1	0.88	32.31
1776	1	1.31	139.56
324	1	3.12	75.00
1634	1	3.06	197.88
1527	1	1.25	99.94
1639	1	1.44	151.00
1792	1	764.75	1193.56 (244.50)
1541	0	1914.50	380.12 (40.00)
1761	0	60.62	2126.75 (452.50)

A track was considered correct if cow left the scene through the correct gate and within 60 s of her RFID-tag registration by the respective gate. In total there were 26 tracks considered, and 23 were correctly tracked, while 3 of the tracks were lost at some point (no longer possible to confirm real-ID), (Figure 17). Note that some of these tracks were quite long and if a track is lost, it is highly unlikely that it will be found again. The longest successfully tracked sequence was 20 minutes long. The three tracks that failed were manually inspected to find the point in time where the error occurred. In one case tracker failed at the border of the image, at the overlap between stitched frames, most likely because cows were more distorted in this area from both viewing angles. Note also that two detections were merged in this overlapping area, after the camera calibration, which includes some errors. The other two cases were a case of ID-shifting due to a densely crowded scenario and confusion due to the earlier made error. Given those 26 starting points, the tracker was able to maintain the correct position in a total of 101.29 minutes or 225 s in average per starting point. Note that these numbers only show the complexity of the dataset. They should not be interpreted as mean time to failure as most of the tracks are not lost entirely but detected at the exit borders of the scene.

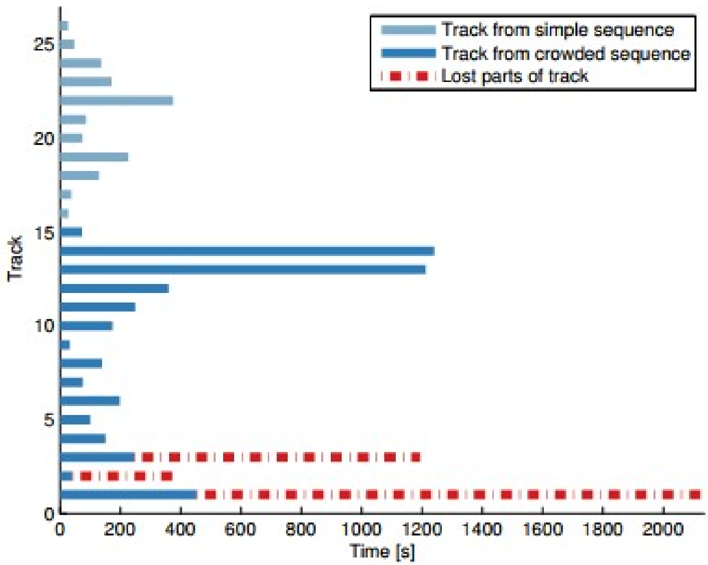

Figure 17. Here all 26 tracks in the dataset are shown in the y-axis and how long the tracker was able to follow each of them in the simple/crowded sequence.

5 Discussion

5.1. Prerequisites for successful implementation of computer vision systems in livestock production

With recent advances in the fields of computer vision and deep learning, as well as affordable computational power, systems based on computer vision could become the solution needed for surveillance of animals in different production systems (Giot, El-Abed and Rosenberg, 2013; Sellers and Hirasaki, 2014; Kulikov et al., 2014; Nilsson et al., 2015; Banhazi and Tscharke, 2016). However, the vast majority of current computer vision systems for monitoring dairy cows are still in the developmental phase and do not provide the flexibility/functionality required for continuous monitoring of animals. Most leading-edge examples on detecting cows by using video cameras have been focused on monitoring areas where the orientation of the cows was known due to physical limitations imposed by the surroundings. Two examples of such work are the Viola-Jones based detector of Porto et al. (2012) for detecting cows at the feed barrier and the work of Martinez-Ortiz et al. (2013) to detect and track cow heads in narrow entrance corridors.

The key-concepts forming the framework needed for robust solutions for automated and accurate tracking/identification of animals, as well as extended behavioural analysis features are not fully established yet. Thus, investigating the opportunities and limitations of recent advances in computer vision and deep learning will facilitate the development of modules capable of monitoring animal health/welfare/behaviour related parameters at low computational cost and in real- or near-real-time manner.

Our research has provided the first step towards the new applications for dairy cows regarding CV's ability to detect interactions between individuals, WatchDog-functionality, tracking capacity, as well as initial function classification.

5.2. Importance of a good mathematical model for describing a cow as an object

As Nasirahmadi et al. (2017) mentioned in their review, successful monitoring of individual animals by computer vision systems is highly dependent on the model describing an animal as a mathematical object. The model should be detailed enough to register behavioural changes, animals' position and movement, as well as have the flexibility for further extension at a low computational cost.

When looking at the overall accuracy of the Behavioural Detector module based on our seven-point shape model, performance and accuracy varied depending on different test-conditions:

- no “line border features”, single frame: accuracy 79.2%,
- with “line border features”, single frame: accuracy 83.1%,
- with “line border features”, three consecutive frames: accuracy 85.1%.

The main difference between these three different test-conditions was the number of geometrical features involved in the analysis as well as a number of frames used for confirming each of the investigated behaviours. This was done to assess the future computational costs, and predict the capacity required for storing the recordings (Paper I).

As could be seen from the confusion matrix, differentiation between similar (based on geometrical shape only) object classes (e.g., Body sniffing and Head pressing) was inconclusive as classifier blended. The idea behind the pre-selected behaviours used for the analysis was to choose something that belonged to entirely different classes from the assessment point of view (positive and negative behaviours) while being very similar shape-vice. This was done to investigate the limitations and fall-backs of the seven-point shape model right from the beginning since the manual annotation with landmarks is a very time-consuming process. The results showed that the proposed model was suitable for representation of complex behaviours and could be applied to different situations while providing the necessary flexibility. The utilisation of additional features and time dimension feature could help in resolving the issue above of blending similar behaviours and provide the potential for further algorithm improvement as well as in decreasing number of false negatives.

While complex social interactions involve such an important parameter as duration, higher classifier accuracy is difficult to achieve only from static landmark points in single frames. One possible solution is to extend the pool of training data (to assure greater variability of behavioural states and reliability of classes in training sequences). This will allow us to re-evaluate the existing features (both from the shape perspective and from the level-of-detail-needed one) to see how the complexity of different behaviours and behavioural states could be described and computed. Depending on different areas of interest in the barn or on different research questions, certain adjustments in the algorithm will be needed to shift between real-time implementation and post-situational analysis (on already recorded data).

Considering the performance of the classifier, similar object classes (2-D representation of behavioural state based on shape model only, a single frame) could be merged to form simplified interaction descriptions (e.g., Head butting and Head pressing into general butting). Such simplified snapshots of performed behaviours could provide quick benchmark (related to cow traffic mainly) of the specific area in the dairy barn to evaluate any adjustments in management practices. One possible (and the most logical) addition to this simplified behavioural overview will be to add heatmap function to the analysis, allowing visualisation of the spatial distribution of cows and time spent either at the specific set of coordinates or proximity to other cow/object of interest.

5.3. The complexity of the approach used for detection of individual animals (Papers III and IV)

General purpose object detection frameworks such as YOLO (Redmon et al., 2016; Redmon and Farhadi, 2017) and SSD (Liu et al., 2016) have outstanding performance. They do, however, focus on detecting objects of varying size and aspect ratio but with a fixed orientation. He et al. (2015) have considered rotated bounding boxes to generate general-purpose objected proposals. They provide a local optimisation over position, size, aspect ratio and rotation from the initial set of sampled windows. In the approach proposed in our work, the size and aspect ratio are known and fixed, and an exhaustive search for position and rotation is performed. That way the risk of not considering relevant rotations is eliminated. Also, the methodology proposed by He et al. (2015) is based on handcrafted features assuming that image borders are unlikely to contain objects and that the object of interest covers a significant part of the interior of the image.

Another way for accurate object detection/classification is to combine segmentation algorithms such as DeepMask (Pinheiro, Collobert and Dollar, 2015) and SharpMask (Pinheiro et al., 2016) with an object detector such as MultiPathNet (Zagoruyko et al., 2016). This results in object detections augmented with pixel level segmentation. From those segmentations, rotated bounding boxes could be generated or, even better, distance measures could be made using the segmentation directly. However, this requires manual pixel level segmentation of training data, while the approach suggested in our work only requires a few manual clicks per training object.

Thus, the second most important task in our work was to understand the limitations of the cow detector itself, as the most crucial component of the developed WatchDog system (Paper III). Most, 94.5%, of the images were perfectly interpreted, i.e. all cows present were detected and no extra detections.

The majority of errors were made in crowded situations where more than two cows were present and detected and thus classified correctly according to the watchdog criteria. Only a single case was found where the WatchDog erroneously discarded a frame, resulting in an overall accuracy of 99.9%. Note that this is a per-frame accuracy and an interaction consists of multiple frames, so missing an entire interaction would be extremely unlikely even though the frames within that interaction would be somewhat correlated. In total, those 1000 images contained 2041 cows. 16 of those cows were not detected, and 48 extra detections were made yielding a cow hit rate of 99.22% with a false alarm rate of 2.35%.

5.4. Inconsistency in the Cow Detector performance during the different seasons

Half a year after the initial recordings were made, another month of video material was collected (see Figure 18), and the original cow detector was tested on that data.

Figure 18. Example frames from a second recording set made half a year after the initial recordings.

It did not perform very well. A set of 931 frames was sampled from the new recording and passed to the detector. It produced annotated images that were then inspected manually. Among those, 510 or 55% of the images were correctly interpreted in the same sense as above, while 421 or 45% contained some mistake. A lot of false detections were made. Especially in overexposed areas and there were also significantly more misses. In a random sample consisting of 50 of the 421 erroneous frames, there was in total 128 cows correctly detected, 28 extra detections made and 45 cows missed.

Two main differences between this new data set and the old one have been identified. First, the cameras have become quite dirty. They were cleaned a few times during the experiments, but they became dirty quite fast, so for a system like this to become useful, it will have to be able to handle somewhat dirty cameras. Second, the sun was low in the sky at the time of year of the new recording. That results in a different lighting situation. It is an indoor scene, but windows are letting in the sunlight.

From this new data set, the 421 frames where the old detector made errors, as well as three frames that contained no errors, were manually annotated in the same way as before. They contained in total 2880 cows. The detector was retrained using both the new and the old data, which resulted in a detector with more smooth performance across the different seasons. A validation set consisting of 10% of the annotated data (196 frames) was separated out and not used during the training. This set was used to evaluate the new detector. In total it contained 408 cows, and 396 of these were detected with only six extra detections. That gives a cow hit rate of 97% with a false alarm rate of 2.9%. Regarding frames, 92.8% were correctly interpreted while there were mistakes made in 14 of the frames. Among those 14 frames, 13 contained more than one cow and thus the hit rate of the watchdog finding frames with two or more cows were 99.5%.

5.5. Tracking of individual cows and method evaluation (Paper IV)

The current state-of-the-art for detecting cows freely moving around the barn was presented by Porto et al. (2013). They used a Viola-Jones based detector and needed six cameras at 4.6 meters height to cover a 15.4 × 3.8 m area to detect cows in three different orientations: vertical, horizontal and diagonal with a hit rate of 90%. They did not use separate detectors for the two different diagonals which means that producing rotated bounding box from their results would not be straightforward. In work presented in this thesis, only three cameras at 3.6 meters height were needed to cover an 18 × 6 m area and detect cows in 32 different orientations with a hit rate of 97%. Note that the hit rates are from different datasets and that the dataset used in this work has larger variations in viewpoints due to the use of fewer cameras at a lower height to cover a larger area.

The Tracker module was developed and tested at the end of the project, being more of a proof-of-concept, since the idea of using the passive data-markers for individual identification of animals was never tested before. Even considering the limited time available for the implementation of the Tracker module, a lot of potentially interesting information was gathered and separated into different classes for further development. The value of non-invasive continuous tracking system capable of identifying the individuals is tremendous and could help in resolving the common overstocking problems of modern dairy barns by assuring the optimal flow of animals and benchmarking-on-the-go.

To test the Tracker module, certain simplifications in the approach were taken. During the study, the exit gates did not register the exit event until the cow had been gone from the scene for a few seconds. Also, one of the entry gates was a place quite far outside of the observed area, which meant that the timings of the exit registrations were more reliable than the timings of the entry registrations. To mitigate the effect of this, the recorded video was reversed in time, and the exit gates were used as entry gates and vice versa. Also, synthetic observations with low probabilities were inserted at the entry and exit states when the actual detection there was lower than the synthetic one. These detections kept the cow tracks in those states during the time between the gate registration, and that enough of the cow appears in the image for a detection to be made. Also, the cows present in the scene at the start of the reversed video were manually marked and given a synthetic id-number. This meant that no exit information was available for these cows. Instead, a different exit criterion was used (for all cows): if a cow's optimal position was one of the synthetic exit gate detections for more than 0.5 s consequentially, it was considered an exit and removed from the tracking. This means that the exit events from the gates were not used by the tracker and could instead be used to evaluate the results.

While considering the average duration of successful tracking events (approximately 225 seconds) and gradually decreasing accuracy in over-crowded scenes, one should bear in mind that the occurrence of errors (false-ID) do not indicate the limitations of the proposed solution. As mentioned previously, the identification error (when Tracker blends the real ID-numbers of cows that are in close proximity to each other) only indicates the per-frame failure. By extending the pool of potential data-collection points, one should be able to recover the initial detection and place the correct ID-marker on the object of interest. Our assumptions suggest that the system will benefit from more cameras installed all over the dairy barn, specifically around areas with selection gates or narrow passages, creating the extended network of passive data-markers.

5.6. Methodological considerations

Due to the limited number of studies in the field of applied computer vision for studying dairy cattle behaviour, it is somewhat difficult to compare pros and cons of different methods/techniques used under different circumstances. Therefore, all the potential pitfalls mentioned in this work apply mostly to our recording setup and may vary from farm to farm, leaving the room for future research and improvements.

One of the primary concerns was the number of cameras used during the studies. As experience showed, using the three cameras with wide (134 degrees) viewing angle resulted in significant overlap between the recorded images. Such overlap could make the detections and analysis more difficult at times, due to obscured/distorted object coordinates, increasing the total computational cost. For all the future experiments, the number of cameras needed for each ROI should be based on the actual scene reconstruction and specific research question and not hardware specification of the camera only.

Another potential add-on to the existing setup is to eventually increase the resolution of recorded video material (step from default 800x600 pixels towards Full HD resolution) as that could increase the precision of detections and add new layers of information. However, with that in mind, the system should be still capable of recording the substantial amounts of data without increasing the storage cost. One possible solution for this could be to divide the range of features for monitoring into immediate (requiring lower resolution due to the simplicity of task) and offline (with higher resolution and additional information). Computer vision solutions with different levels of data-analysis-and-visualisation will also require the efficient infrastructure for data exchange as well as the ability to communicate with other existing data systems installed on-site.

6 Conclusions

The first study introduced a new approach for the detection of social interactions in dairy cattle. This could be considered as the first step in the development of an entirely automatic learning-based framework for behaviour monitoring, showing nearly 85% of accuracy when comparing it to ground truth levels confirmed by an experienced observer.

The second study concluded that the Floor/Claw module showed the potential for further development and could be used as a tool for practical assessment of actual weight load/animal distribution in areas of interest. In both cases (160 mm and 560 mm concrete slats), it could be concluded that the floor closest to the entrances to AMS will be slightly more loaded.

The third study resulted in the development of a CNN-based cow detection system. The proposed solution can detect and count the cows present in the image with high precision. 92.8% of the test images were perfectly interpreted in the sense that the system was able to place a rotated rectangle on each cow and nowhere else. This detector was used to discard 50% of the recorded video as uninteresting while only losing 4% of the interesting video. If detecting a single frame per interaction is enough, 60% of the test dataset could be automatically discarded without losing any of the interactions. Regarding detection of single cows, the hit rate was 97% with a false alarm rate of 2.9%. Note that these numbers depend on how significant portion of the recorded video is interesting and how often the scene is crowded (which is when most mistakes are made). This will vary between different farms and studies.
Another important conclusion was that the even though it is an indoor scene, the lighting variations caused by the different seasons are significant. It was not enough to collect data from three months. Instead, both training data and evaluation data from different seasons are needed to build a system that can operate all year around.

The fourth and final study investigated and proposed the flexible and non-invasive computer vision system for tracking and identification of individual cows. The cows and their real-id numbers were tracked in a waiting area before automatic milking stations. The system was investigated on a real conventional farm with all the real-world issues, such as over year illumination changes and spider webs obscuring the field of view of the cameras. The proposed system is a crucial stepping stone towards a fully automated tool for continuous monitoring of cows and their interactions with other individuals and the farm-building environment. Furthermore, the system is based on several state-of-the-art deep learning methods, which enabled handling several real-world issues. Experiments indicate that a cow could be tracked close to 4 minutes before failure cases emerge and that cows could be successfully tracked for over 20 minutes in mildly-crowded (<10 cows) scenes.

7 Future research plans

Future work aims to improve the performance and flexibility of the modules from the WatchDog system as well the development of new building blocks or modules which will extend the functionality of the proposed solution by:

- source code optimisation to achieve the opportunity to use WatchDog on different platforms and embedded hardware at lower computational cost without losing the functionality;
- adding the additional features to Behavioural Detector module that will allow spatial analysis as well as social network analysis based on proximity between the individuals. This will allow the studies investigating hierarchical order within the herd, maternal relationship and preference tests (both towards other individuals and specific areas of interest);
- extending the range of functions for the Behavioural Detector module by adding features that will allow calving detection, simple gait recognition and benchmarking of scene complexity;
- investigating the opportunities to create a smartphone-based application with Augmented Reality (AR) functionality for behavioural evaluation on-site (as well as some lameness detection based on back ridge angle);
- making a “For Research Purposes” only graphical user interface (GUI) for WatchDog module responsible for video filtering, allowing more efficient work with the recordings;
- investigating the opportunity for adjusting the developed shape-model to include other animal species into the algorithm;

References

Adamczyk, K., Górecka-Bruzda, A., Nowicki, J., Gumułka, M., Molik, E., Schwarz, T., and Klocek, C. (2015). Perception of the environment in farm animals. A review. *Annals of Animal Science, 15*, pp. 565 – 589.

Ahrendt, P., Gregersen, T. and Karstoft, H. (2011). Development of a real-time computer vision system for tracking loose-housed pigs, *Computers and Electronics in Agriculture*, 76 (2), pp. 169-174.

Ardo, H., Guzhva, O., Nilsson, M. and Herlin, A.H. (2017). A CNN-based cow interaction watchdog. *IET Computer Vision*, pp. 1-7.

Awad, A.I. (2016). From classical methods to animal biometrics: A review on cattle identification and tracking, *Computers and Electronics in Agriculture*, 123.

Banhazi, T.M. and Tscharke, M. (2016). A brief review of the application of machine vision in livestock behaviour analysis, *Journal of Agricultural Informatics*, 7 (1).

Barkema, H. W., Y. H. Schukken, T. J. G. M. Lam, M. L. Beiboer, G. Benedictus and A. Brand. (1998). Management practices associated with low, medium, and high somatic cell counts in bulk tank milk. *Journal of Dairy Science, 81*, pp.1917-1927.

Barkema, H., von Keyserlingk, M., Kastelic, J., Lam,T., Luby, C., Roy, J.-P., LeBlanc, M., Keefe, G. and Kelton, D. (2015). Invited review: Changes in the dairy industry affecting dairy cattle health and welfare, *Journal of Dairy Science 98 (11)*, pp. 7426-7445.

Bergsten, C., M. Jansson Mörk, M. Pringle, and Y. Persson. (2017). Can Interdigital dermatitis (Foot rot) be treated without antibiotics? Pages 313-314 in Proceedings of the 19th International Symposium and 11th Conference Lameness in Ruminants A. Fiedler and K. Schindhelm, ed, Munich.

Bookstein, F.L. (1997). Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. *Med. Image Anal. 1*, pp. 225-243.

Boyland, N. K., Mlynski, D. T., James, R., Brent, L. J., & Croft, D. P. (2016). The social network structure of a dynamic group of dairy cows: From individual to group level patterns. *Applied Animal Behaviour Science, 174*, pp. 1 – 10.

Busse, M., Schwerdtner, W., Siebert, R., Doernberg, A., Kuntosch, A., König, B., Bokelmann, W. (2015). Analysis of animal monitoring technologies in Germany from an innovation system perspective. *Agric. Syst.* 138, pp. 55–65.

Cangar, O., Leroy, T., Guarino, M., Vranken, E., Fallon, R., Lenehan, J., Mee, J. and Berckmans, D. (2008). Automatic real-time monitoring of locomotion and posture behaviour of pregnant cows prior to calving using online image analysis, *Computers and Electronics in Agriculture* 64 (1), pp. 53-60.

Carne, S., Caja, G., Ghirardi, J. and Salama, A. (2009). Long-term performance of visual and electronic identification devices in dairy goats, *Journal of Dairy Science*, 92 (4), pp. 1500-1511.

Castro, I.M., Gygax, L., Wechsler, B. and Hauser, R. (2012). Effect of short and long periods of separation on agonistic behaviour, injuries and stress in Herens cows kept in loose housing, *Applied Animal Behaviour Science 136 (2-4), pp. 96-103.*

Coulon, M., Baudoin, C. & Heyman, Y., & Deputte, B.L. (2011). Cattle discriminate between familiar and unfamiliar conspecifics by using only head visual cues. *Animal Cognition, 14,* pp. 279 – 290.

Cook, N. B., T. B. Bennett, and K. V. Nordlund. (2004). Effect of free stall surface on daily activity patterns in dairy cows with relevance to lameness prevalence. *J. Dairy Sci., 87*, pp. 2912–2922.

Decker, J. E., McKay, S. D., Rolf, M. M., Kim, J., Molina Alcalá, A, Sonstegard, T. S., Hanotte, O., Götherström, A., Seabury, C. M., and Praharani, L. (2014). Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS Genetics 10, e1004254.*

DeVries, T. J., von Keyserlingk, M. A. G. and Weary, D. M. (2004). Effect of feeding space on the inter-cow distance. Aggression, and Feeding Behaviour of Free-Stall Housed Lactating Dairy Cows. *J. Dairy Sci.,* 87.

Dominiak, K. and Kristensen, A. (2017). Prioritizing alarms from sensor-based detection models in livestock production - A review on model performance and alarm reducing methods, *Computers and Electronics in Agriculture*, 133, pp. 46-67.

Dziuk, P. (2003). Positive, accurate animal identification, *Animal Reproduction Science*, 79 (3-4), pp. 319-323.

Geers, R. and Madec, F. (2006). Livestock production and society, Wageningen Academic Publishers, The Netherlands, 2006.

Giot, R., El-Abed, M. and Rosenberger, C. (2013). Fast computation of the performance evaluation of biometric systems: Application to multibiometrics, *Future Generation Computer Systems*, 29 (3), pp. 788-799.

Gulshan, V., Rother, C., Criminisi, A., Blake, A., Zisserman, A. (2010). Geodesic star convexity for interactive image segmentation. In: Proc. of the IEEE Conference on Computer Vision and Pattern Recognition San Francisco, USA, pp. 3129–3136.

Guzhva, O., Ardö, H., Herlin, A., Nilsson, M., Åström, K. and Bergsten, C. (2016). 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. *Computers and Electronics in Agriculture*, 127, pp. 506–509.

Grandin, T. (2015). *Improving animal welfare: A practical approach (2nd ed). Oxfordshire, UK: CABI.*

Haidet, K.K., Tate, J., Divirgilio-Thomas, D., Kolanowski, A. and Happ, M.B. (2009). Methods to Improve Reliability of Video Recorded Behavioral Data. *Res. Nurs. Health.,* 32(4), pp. 465-474.

Halachmi, I., Metz, J.H.M., Maltz, E., Dijkhuizen, A.A., Speelman, L. (2000). Designing the optimal robotic milking bar, Part 1: Quantifying facility usage. J. *Agric. Eng. Resour.* 76, pp. 37–49.

Hartley, R. I. and Zisserman, A. (2004). Multiple View Geometry in Computer Vision. Cambridge University Press, ISBN: 0521540518, second edition, 2004.

Hartley, R.I. and Zisserman, A. (2004). Multiple View Geometry in *Computer Vision, 2nd Edition*, Cambridge University Press.

He, K. and Lau, R.W. (2015). Oriented object proposals, in: *The IEEE International Conference on Computer Vision (ICCV)*, 2015.

Hedlund, L., & Løvlie, H. (2015). Personality and production: Nervous cows produce less milk. *Journal of Dairy Science, 98*, pp. 5819 – 5828.

Hemsworth, P. (2003). Human-animal interactions in livestock production. *Appl. Anim. Behav. Sci*, 81.

Herlin, A.H. and Frank, B. (2007). Effects of protective gates at concentrate feed stations on behaviour and production in dairy cows: A brief note, *Applied Animal Behaviour Science 103 (1-2), pp. 167-173.*

Hermans, G., Ipema, A., Stefanowska, J. and Metz, J. (2003). The Effect of Two Traffic Situations on the Behaviour and Performance of Cows in an Automatic Milking System, *Journal of Dairy Science 86 (6), pp. 1997-2004.*

Horn, B.J.P. (2000). Tsai's Camera Calibration Method Revisited. <http://people.csail.mit.edu/bkph/articles/Tsai_Revisited.pdf>.

Ioffe, S. and Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift., arXiv:1502.03167v3

Jabbar, K.A., Hansen, M.F., Smith, M. and Smith, L. (2016). Locomotion traits of dairy cows from overhead three-dimensional video. *Visual observation and analysis of Vertebrate And Insect Behavior*, 2016.

Kaiming, H., Zhang, X., Ren, S. and Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. CoRR, abs/1502.01852.

Kashiha, M., Pluk, A., Bahr, C., Vranken, E., Berkmans, D. (2013). Development of an early warning system for a broiler house using computer vision. *Biosyst. Eng.* 116, pp. 36–45.

Kiley-Worthingthon, M. & De La Plain, D. (1983). The social organization of the herd. In Fölsch, D. W. (Ed.), The behaviour of beef suckler cattle (Bos taurus) (pp. 105 – 123). Zürich, Switzerland: Verlag Basel.

Kilgour, R. J. (2012). In pursuit of 'normal': A review of the behaviour of cattle at pasture. *Appl. Anim. Behav. Sci*, 138.

Kondo, S., Sekine, J., Okubo, M., & Asahida, Y. (1989). The effects of group size and space allowance on the agonistic and spacing behaviour of cattle. *Applied Animal Behaviour Science, 24,* pp. 127 – 135.

Kremer, P. V., S. Nueske, A. M. Scholz, and M. Foerster. (2007). Comparison of claw health and milk yield in dairy cows on elastic or concrete flooring. *J. Dairy Sci., 90,* pp. 4603–4611.

Kulikov, V.A., Khotskin, N.V., Nikitin, S.V., Lankin, V.S., Kulikov, A.V. and Trapezov, O.V. (2014). Application of 3-D imaging sensor for tracking minipigs in the open field test, *Journal of Neuroscience Methods*, 235, pp. 219-225.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., and Berg, A.C. (2016). SSD: Single shot multibox detector, in: *European Conference on Computer Vision, Springer*, 2016, pp. 21-37.

MacKay, J. R. D., Haskell, M. J., Deag, J. M., & van Reenan, K. (2014). Fear responses to novelty in testing environments are related to day-to-day activity in the home environment in dairy cattle. *Applied Animal Behaviour Science, 152*, pp. 7-16.

Magnusson, M., Ventorp M., Nilsson C. (2008). Drainage Capacity of Concrete Slatted Floors for Dairy Cattle. CIGR, Available: http://www.cigrjournal.org/index.php/Ejounral/article/view/1029/1022# [2016-05-06].

Marchant, J.A. (1988a). Computer vision in agricultural engineering. *Agric. Eng., 43(2)*, pp. 40-42.

Marino, L. and Allen, K. (2017). The Psychology of Cows. *ABC 4 (4)*, pp. 474-498.

Martin, P. and Bateson, P. (2007). Measuring Behaviour: An Introductory Guide. Cambridge University Press, Cambridge, 2007.

Martinez-Ortiz, C.A., Everson, R.M. and Mottram, T. (2013). Video tracking of dairy cows for assessing mobility scores.

McTavish, E. J., Decker, J. E., Schnabel, R. D., Taylor, J. F., & Hillis, D. M. (2013). New World cattle show ancestry from multiple independent domestication events. *Proceedings of the National Academy of Sciences, 110,* E1398 – E1406.

Melin, M., Hermans, G., Pettersson, G., Wiktorsson, H. (2006). Cow traffic in relation to social rank and motivation of cows in an automatic milking system with control gates and an open waiting area. *Appl. Animal Behav. Sci.* 96, pp. 201–214.

Melin, M., Svennersten-Sjaunja, K., Wiktorsson, H. (2005). Feeding patterns and performance of cows in controlled cow traffic in automatic milking systems. *J. Dairy Sci.* 88, pp. 3913–3922.

Miguel-Pacheco, G., Kaler, J., Remnant, J., Cheyne, L., Abbort, C., French, A., Pridmore, T., Huxley, J. (2014). Behavioural changes in dairy cows with lameness in an automatic milking system. *Appl. Animal Behav. Sci.* 150, pp. 1–8.

Nadimi, E., Jørgensen, R., Blanes-Vidal, V. and Christensen, S. (2012). Monitoring and classifying animal behavior using ZigBee-based mobile ad hoc wireless sensor networks and artificial neural networks, *Computers and Electronics in Agriculture*, 82, pp. 44-54.

Nasirahmadi, A., Edwards, S.A. and Sturm, B. (2017). Implementation of machine vision for detecting behaviour of cattle and pigs. *Livestock Science 202*, pp. 25-38.

Nilsson, M., Herlin, A.H., Ardö, H., Guzhva, O., Åström, K., Bergsten, C. (2015). Development of automatic surveillance of animal behaviour and welfare using image analysis and machine learned segmentation technique. *Animal* 9, pp. 1859–1865.

Oczak, M., Ismayilova, G., Costa, A., Viazzi, S., Sonoda, L., Fels, M., Bahr, C., Hartung, J., Guarino, M., Berckmans, D., Vranken, E. (2013). Analysis of aggressive behaviours of pigs by automatic video recordings. *Comput. Electron. Agric.* 99, pp. 209–217.

Olofsson, J. (1999). Competition for total mixed diets fed for ad libitum intake using one or four cows per feeding station. *J. Dairy Sci.* 82, pp. 69–79.

Otsu, N. (1979). A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybernetics. 9, pp. 62–66.

Pinheiro, P.H.O., Collobert, R. and Dollár, P. (2015). Learning to segment object candidates. CoRR, abs/1506.06204.

Pinheiro, P.H.O., Lin, T-Y., Collobert, R. and Dollár, P. (2016). Learning to refine object segments. CoRR, abs/1603.08695.

Phillips, C. (2002). *Cattle behavior and welfare. Oxford, UK: Blackwell Publishing.*

Platz, S., F. Ahrens, J. Bendel, H. H. D. Meyer, and M. H. Erhard. (2008). What happens with cow behavior when replacing concrete slatted floor by rubber coating: A case study. *J. Dairy Sci., 91,* pp. 999–1004.

Proudfoot, K. L., Weary, D. M., & von Keyserlingk, M. A. (2012). Linking the social environment to illness in farm animals. *Applied Animal Behaviour Science, 138*, pp. 203 – 215.

Polikarpus, A., Kaart, T., Mootse, H., Arney G. (2015). Influences of various factors on cows' entrance order into the milking parlour. *Appl. Anim. Behav. Sci*., 166.

Porto, S. MC., Arcidiacono, C., Anguzza, U. and Cascone, G. (2015). The automatic detection of dairy cow feeding and standing behaviours in free-stall barns by a computer vision-based system. *Biosystems Engineering*, 133, pp. 46–55.

Porto, S.M., Arcidiacono, C., Anguzza, U. and Cascone, G. (2013). A computer vision-based system for the automatic detection of lying behaviour of dairy cows in free-stall barns, *Biosystems Engineering*, 115 (2).

Porto, S.M.C., Arcidiacono, C., Anguzza U. and Cascone, G. (2012). Detecting cows at the feed barrier by means of an image analysis algorithm. International Conference RAGUSA SHWA 2012, September 3-6, 2012, Ragusa - Italy.

Prescott, N.B., Mottram, T.T., Webster, A.J.F. (1998). Relative motivations of dairy cows to be milked or fed in a Y-maze and an automatic milking system. *Appl. Animal Behav. Sci.* 57, pp. 23–33.

Rault, J-L. (2012). Friends with benefits: Social support and its relevance for farm animal welfare. *Applied Animal Behaviour Science, 136*, pp. 1 – 14.

Redmon, J. and Farhadi, A. (2017). Yolo9000: Better, faster, stronger. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.

Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016.

Ren, E., He, K., Girshick, R. and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks, in: *C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, R. Garnett (Eds.), Advances in Neural Information Processing Systems 28, Curran Associates, Inc., 2015*, pp. 91-99.

Rousing, T., Wemelsfelder, F. (2006). Qualitative assessment of social behaviour of dairy cows housed in loose housing systems. *Appl. Animal Behav. Sci.* 101, pp. 40–53.

Rumelhart, D.E., Hinton, G.E., Williams, R.J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 6088, 533–536., doi:10.1038/323533a0.

Rutten, C., Velthuis, A., Steeneveld, W., and Hogeveen, H. (2013). Invited review: Sensors to support health management on dairy farms. *Journal of Dairy Science* 96 (4), pp. 1928-1952.

Schreiner, D. A. and P. L. Ruegg. (2003). Relationship between udder and leg hygiene scores and subclinical mastitis. *Journal of Dairy Science, 86*, pp. 3460-3465.

Sellers, W.I. and Hirasaki, E. (2014). Markerless 3D motion capture for animal locomotion studies, *Biology Open*, 3 (7), pp. 656-668.

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R. and Lecun, Y. (2014). Overfeat: Integrated recognition, localization and detection using convolutional networks.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition, *CoRR* abs/1409.1556.20435.

Simonyan, K. and Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. Published as a conference paper at ICLR 2015, arXiv:1409.1556.

SIS, Swedish Standards institute. (2007). Precast concrete products – Floor slats for livestock. Stockholm, SIS Förlag AB (SS-EN 12737:2004+A1:2007).

Shin, D., Kang, H., & Seo, S. (2017). Social relationships enhance the time spent eating and intake of a novel diet in pregnant Hanwoo (Bos taurus coreanae) heifers. *PeerJ, 5,* e3329

Sveberg, G., Refsdal, A., Erhard, H., Kommisrud, E., Aldrin, M., Tvete, I., Buckley, F., Waldmann, A., Ropstad, E. (2013). Sexually active groups in cattle – a novel estrus sign. *J. Dairy Sci.* 96, pp. 4375–4386.

Taylor, A. A., and Davis, H. (1998). Individual humans as discriminative stimuli for cattle Bos taurus. *Applied Animal Behavior Science, 58*, pp. 13 – 21.

Telezhenko, E. and C. Bergsten. (2005). Influence of floor type on the locomotion of dairy cows. *Applied Animal Behaviour Science, 93,* pp. 183-197.

Telezhenko, E. (2007). Effect of flooring system on locomotion comfort in dairy cows: Aspects of gait, preference and claw condition. *PhD Thesis. Swedish University of Agricultural Sciences (SLU), Skara, Sweden.*

Tsai, D.-M., Huang, C.-Y., 2014. A motion and image analysis method for automatic detection of estrus and mating behaviour in cattle. *Comput. Electron. Agric.* 104, pp. 25–31.

Tsai, R.Y. (1986). An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Miami Beach, FL, USA, pp. 364-374.

Tullo, E., Fontana, I., Gottardo, D., Sloth, K. and Guarino, M. (2016). Technical note: Validation of a commercial system for the continuous and automated monitoring of dairy cow activity, *Journal of Dairy Science*, 99 (9), pp. 7489-7494.

Van der Stuyft, E., Schofield, C.P., Randall, J.M., Wambacq, P. and Goedseels, V. (1991). Development and application of computer vision systems for use in livestock production. *Computers and Electronics in Agriculture, 6,* pp. 243-265.

Viola, P. and Jones. M. (2004). Robust Real-time Object Detection. *International Journal of Computer Vision, 57(2)*, pp 137 – 154.

Vieira, A. D. P., de Passillé, A. M., & Weary, D. M. (2012). Effects of the early social environment on behavioral responses of dairy calves to novel events. *Journal of Dairy Science, 95,* pp. 5149 – 5155.

Vokey, F. J., C. L. Guard, H. N. Erb, and D. M. Galton. (2001). Effects of alley and stall surfaces on indices of claw and leg health in dairy cattle housed in a free-stall barn. *J. Dairy Sci., 84*, pp. 2686–2699.

von Keyserlingk, M., Rushen, J., de Passille, A. and Weary, D. (2009). Invited review: The welfare of dairy cattle. Key concepts and the role of science, *Journal of Dairy Science 92 (9), pp. 4101-4111.*

Wathes C., (2007). Precision livestock farming for animal health, welfare and production. *In*: Aland A., ed. *Proceedings of the 13th International congress in animal hygiene, animal health and welfare, biosecurity, environmental protection, sustainability of animal husbandry, June 17-21, 2007, Tartu, Estonia*, pp. 397-404.

Young, G. and Householder, A.S. (1938). Discussion of a set of points in terms of their mutual distances. *Psychometrika*, 3(1), pp. 19–22.

Zagoruyko, S., Lerer, A., Lin, T-Y., Pinheiro, P.H.O., Gross, S., Chintala, S. and Dollár, P. (2016). A multipath network for object detection. CoRR, abs/1604.02135, 2016.

Populärvetenskaplig sammanfattning

I mjölkproduktionen blir djurbesättningarna med lösgående djur allt större och arbetet effektiviseras genom mer mekanisering och automatisering. Skötseltiden för varje enskilt djur blir kortare samtidigt som det ställs stora krav på djurens tillsyn för deras välfärd och hälsa. Under senare år har flera tekniska hjälpmedel (sensorer) införts i lösdriftsstallar som gör det möjligt att förbättra övervakningen av djurs hälsa, fruktsamhet och produktion.

Användning av kameror, där bilderna analyseras av en dator, finns redan utvecklat för att upptäcka juversjukdom, hälta och bedöma hull hos kor, samt för att se var djuren befinner sig i stallet. Genom sådan bildanalys är det också möjligt att registrera vissa beteenden på individnivå hos nötkreatur. För att förstå hur korna fungerar som grupp och på individnivå görs idag manuella beteendestudier som är mycket tidskrävande och därmed kostsamma. En målsättning är därför att utveckla teknologi och verktyg för att automatiskt detektera sådana sociala interaktioner på ett tillförlitligt sätt.

Syftet med detta arbete var att undersöka möjligheter och begränsningar för att använda bildanalys för att studera kors beteende, rörelser, positioner och samspel dem emellan. I och med detta skulle ett första steg kunna tas mot ett helautomatiskt system för kontinuerlig övervakning av djurhälsa och djurvälfärd i modern mjölkproduktion. Vårt första mål var att undersöka om det var möjligt att från en videofilm automatiskt definiera vad som är en ko i hennes stallmiljö samt bestämma hennes position. Därefter fortsatte arbetet att automatiskt bestämma kornas beteende och sociala samspel (interaktioner) med varandra. Ett ytterligare mål var att utnyttja bildanalys för att bestämma klövarnas placering på bärande element av spaltgolv. En frågeställning var hur man automatiskt selekterar intressanta videosekvenser och slutligen hur man kan koppla kor som man ser i bild till dess rätta identitet och sedan följa dem runt i stallet.

I den första studien utvecklades en ny matematisk modell för att beskriva en ko som ett geometriskt objekt. Resultaten visade att den nya modellen, med utnyttjande av maskininlärning, gav möjlighet att studera olika interaktioner mellan korna. Den utvecklade automatiska detektionen fångade upp nästan 85% av interaktionerna jämfört med de som upptäcktes av en erfaren observatör.

I den andra studien vidareutvecklades ett neuronnät (CNN), som är en självlärande algoritm som försöker efterlikna funktionen i biologiska neuronnät (exempelvis hjärnan). Med denna CNN metodik kunde korna och deras klövplaceringar på ett spaltgolv i en samlingsfålla kartläggas. Dessa resultat kan användas som ett verktyg för att bedöma den faktiska belastningen på golvet och därmed för att dimensionera golvet så att det håller för den aktuella kobeläggningen.

I den tredje studien utvecklades en metod för att med hög precision selektera fram intressant videomaterial för beteendestudier, detta genom upptäcka och räkna de kor som fanns i bilden. Nästan 93% av testbilderna tolkades fullständigt i den meningen att systemet kunde placera en rektangel på varje ko och ingen annanstans. Vidare användes denna teknologi för att plocka bort 50% av de inspelade videosekvenserna som inte var intressanta, utan att riskera förlora mer än 4% av de intressanta bilderna. Därmed kunde analyserna effektiviseras och mycket forskningstid sparas.

I den fjärde och sista studien undersöktes hur man kan spåra och identifiera enskilda kor med videokameror och datamarkörer (till exempel selektionsgrindar eller VMS-stationer) som redan finns i ladugårdens datasystem, dvs. utan att sätta fast några sensorer på korna. Systemet fungerade trots att ljuset varierade under året och att spindelnät fanns i taket och skymde kameralinsen till viss del. Studien visade vidare att kor kunde spåras i över 20 minuter (i situationer med <10 kor).

Avhandlingen visade att:

- Det är möjligt att använda bildanalys för identifiering av kor i deras stallmiljö samt att registrera position, aktivitet och interaktioner;
- Med bildanalys kan djuren övervakas kontinuerligt och automatiskt utan att någon apparatur fästs på djuren;
- Med bildanalysteknologi kan man fånga upp förändringar i djurgruppens beteende, som signaler på djurens välbefinnande och hälsa och därmed åtgärda problem tidigt;
- Vid beteendestudier kan bildanalys spara tid och resurser jämfört manuell analys av videoinspelningar;

Popular science summary

The average farm size in dairy production is continuously increasing resulting in increasing mechanisation and automation of on-farm procedures for achieving more efficient workflow. It also results in decreased time for farmers or animal caretakers to spend on observing individual animals. This, combined with the strong public demand for improved animal welfare and health creates demand for new tools (sensors) and methods that could help in monitoring animals´ health, production, and fertility parameters.

The use of cameras and computer vision for automatic analysis of recorded images/video material already resulted in some products aiming to diagnose udder problems, lameness, body condition or position of animals in the barn. The image analysis also gives an opportunity to record certain behaviours of individual cows. To understand how the cows function in a production environment and what kind of relationships they have with other cows, scientists working with animal behaviour use manual observations. These observations are very time consuming and therefore quite expensive. The idea is to replace the manual observations with automatic camera-based tools capable of recognising and analysing social interactions in dairy cows.

The aim of this thesis was to investigate the opportunities and limitations of image analysis for studying cows' behaviours, movements, positions and interactions with other individuals. This development could be the stepping stone towards the fully automatic surveillance system for continuous monitoring of animal health and welfare in modern dairy production.

The first research goal for us was to investigate the possibility to differentiate between the cow and her environment from the recorded video material as well as to find her position using image analysis.

The mathematical model describing the cow from the geometrical perspective was developed and tested. The results showed that the proposed model in combination with some machine learning approach could distinguish between different social interactions in dairy cows. The behavioural detector based on this geometrical model was able to correctly identify 85% of all the behaviours in recordings when compared to an experienced human observer.

The continuation of that work was to add the opportunity to record and differentiate between different social interactions occurring between cows in the dairy barn. The second study was concentrated on developing the artificial neural network (CNN), which is a self-learning algorithm mimicking the work of real mammal brain. This algorithm was developed to detect the positions and placements of cows' claws on concrete floor elements and in that way evaluate the weight load per each of those elements. The potential outcome of such detections is an ability to map cows' positions in areas of interest and evaluate the spatial distribution of animals as well as the pressure they create on floor elements.

The outcome of the third study was a method allowing fully automatic filtering of the recorded video material based on user-defined parameters (number of cows, the distance between them or certain interactions). The, so called, WatchDog algorithm was able to correctly identify almost 93% of all cows in every analysed image resulting in filtering out 50% of video material as irrelevant while potentially losing only around 4% of image data. This approach could save the expert time, previously used for manual processing of the recorded videos.

The fourth and final study investigated the opportunity of continuous tracking and identification of cows in an automatic and non-invasive manner (using only cameras, without the need to put additional sensors on animals). The advantage of the proposed method is that the identification is based on already existing data markers present in management system at any AMS farm (data from selection gates, AMS-stations, automatic feeders etc.). The system showed stable performance even under varying lighting conditions and during different seasons, nevertheless spider webs and dirt on camera lenses. The study showed that the cows could be tracked in bouts of 20 minutes (for mildly-crowded scenes with 10 or less cows).

The thesis provided a solid basis for the further development of automated computer vision systems for monitoring different aspects of animal behaviour and health in modern dairy production.

Acknowledgements

The studies in this thesis were funded by **FORMAS (and all the taxpayers)**; thank you for making this possible!

The long and bumpy road which resulted in me finishing this thesis would not have been possible without the support of a large number of people. I cannot describe the appreciation I feel for all the help and “positive experiences” I received and, therefore, would like to thank you all!

Anders and **Christer** for giving me the freedom to carry out this research in exactly the way I wanted it to be performed, for trusting and supporting my ideas and choices.

Lena, **Charlotte**, and **Kalle** for valuable comments, discussions and all the good questions during our meetings.

Håkan and **Mikael** for the tremendous amounts of patience towards all my questions and also for making me fall in love with the exciting new field of computer vision (and a couple of programming languages along the way).

Linda for being the BEST boss and co-supervisor EVER. And for all the time, trust, and OPPORTUNITIES for personal development within the group and department.

Andrus for all the countless reminders about “go get some sleep already” when seeing me online at night. And, for all the Classic Rock gems you sent me during these four years of my PhD studies. Those songs kept me going even when the energy was low.

Madde and **Kristina** for giving me the opportunity to teach and to evaluate my knowledge and experience by trying to transfer the “essence” to students (and improving A LOT while doing that).

Anne-Charlotte for being the kind and supportive person, colleague, friend.

Lars for being my best friend, and simply by hearing me out in times of need. And also by reminding me about the importance of “training the physical body which will then host the healthier mind”.

Coffee for being the “Juice of Life” and “The Greatest Scientific Inspirer”.

And last (but not least), **Isabel,** my wonderful wife, for being my better half and for being THE MOST UNDERSTANDING person in the world. Your support and love helped me during the hardest of times and guided me across the deepest of research seas.

Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2018:33

This thesis provides knowledge about opportunities and limitations of computer vision and image analysis algorithms for studying cows' behaviours, movements, positions and interactions with other individuals. This development could be the stepping stone towards the fully automatic surveillance system for monitoring of animal health and welfare in modern dairy production.

Oleksiy Guzhva, received his postgraduate education at the Department of Biosystems and Technology, SLU, Alnarp. He received his Doctor of Veterinary Medicine degree at Poltava State Agrarian Academy, Ukraine.

Acta Universitatis Agriculturae Sueciae presents doctoral theses from the Swedish University of Agricultural Sciences (SLU).

SLU generates knowledge for the sustainable use of biological natural resources. Research, education, extension, as well as environmental monitoring and assessment are used to achieve this goal.

Online publication of thesis summary: http://pub.epsilon.slu.se/

ISSN 1652-6880

ISBN (print version) 978-91-7760-206-4

ISBN (electronic version) 978-91-7760-207-1