CORF3D contour maps with application to Holstein cattle recognition using RGB and thermal images

Bhole, Amey; S. Udmale, Sandeep; Falzon, Owen; Azzopardi, George

Published in:
Expert systems with applications

DOI:
10.1016/j.eswa.2021.116354

Expert Systems With Applications 192 (2022) 116354

Contents lists available at ScienceDirect

Expert Systems With Applications

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

CORF3D contour maps with application to Holstein cattle recognition from RGB and thermal images

Amey Bhole [a], Sandeep S. Udmale [b], Owen Falzon [c], George Azzopardi [a,*]

[a] Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen, Nijenborgh 9, 9747AG, Groningen, The Netherlands
[b] Department of Computer Engineering and Information Technology, Veermata Jijabai Technological Institute (VJTI), Mumbai 400019, Maharashtra, India
[c] Centre for Biomedical Cybernetics, University of Malta, Msida MSD2080, Malta

ARTICLE INFO

Keywords:
Animal biometrics
Cattle recognition
Contour detection
ConvNets
Push–pull inhibition
Thermal images

ABSTRACT

Livestock management involves the monitoring of farm animals by tracking certain physiological and phenotypical characteristics over time. In the dairy industry, for instance, cattle are typically equipped with RFID ear tags. The corresponding data (e.g. milk properties) can then be automatically assigned to the respective cow when they enter the milking station. In order to move towards a more scalable, affordable, and welfare-friendly approach, automatic non-invasive solutions are more desirable. Thus, a non-invasive approach is proposed in this paper for the automatic identification of individual Holstein cattle from the side view while exiting a milking station. It considers input images from a thermal-RGB camera. The thermal images are used to delineate the cow from the background. Subsequently, any occluding rods from the milking station are removed and inpainted with the fast marching algorithm. Then, it extracts the RGB map of the segmented cattle along with a novel CORF3D contour map. The latter contains three contour maps extracted by the Combination of Receptive Fields (CORF) model with different strengths of push–pull inhibition. This mechanism suppresses noise in the form of grain type texture. The effectiveness of the proposed approach is demonstrated by means of experiments using a 5-fold and a leave-one day-out cross-validation on a new data set of 3694 images of 383 cows collected from the Dairy Campus in Leeuwarden (the Netherlands) over 9 days. In particular, when combining RGB and CORF3D maps by late fusion, an average accuracy of 99.64% (±0.13) was obtained for the 5-fold cross validation and 99.71% (±0.31) for the leave-one day-out experiment. The two maps were combined by first learning two ConvNet classification models, one for each type of map. The feature vectors in the two FC layers obtained from training images were then concatenated and used to learn a linear SVM classification model. In principle, the proposed approach with the novel CORF3D contour maps is suitable for various image classification applications, especially where grain type texture is a confounding variable.

1. Introduction

The combined effect of intelligent methods and advancement in various technologies has in recent years allowed for the application of these developments in the field of agriculture. This concept known as smart farming has significantly impacted different agricultural sectors (Kamilaris, & Prenafeta-Boldú, 2018), namely organic farming, agribusiness, precision farming, and animal husbandry (Alsahaf, Azzopardi, Ducro, Hanenberg, Veerkamp et al., 2018, 2019; Kumar, & Singh, 2017; Singh, Singh et al., 2020).

Moreover, the rise in the global population has led to a direct impact on food requirements. This condition, coupled with other essential parameters, such as nutritional quality, cost, health, and protection of natural ecosystems with sustainable farming procedures, has driven researchers' and practitioners' efforts to preserve food production standards. As a consequence, it is essential to focus on the fundamental issues associated with the agriculture industry, including productivity, sustainability, environmental impact, and animal welfare concerns. Thus, there is a need to continuously observe, monitor, and analyze the crops as well as animals, with the help of intelligent methods, in order to mitigate these problems. As a result, automatic individual recognition of animals plays a vital role in animal husbandry, where animals are reared for meat, milk, eggs, and other products.

The code (and data) in this article has been certified as Reproducible by Code Ocean: (https://codeocean.com/). More information on the Reproducibility Badge Initiative is available at https://www.elsevier.com/physical-sciences-and-engineering/computer-science/journals.

* Corresponding author.
E-mail addresses: amey.bhole77@gmail.com (A. Bhole), ssudmale@it.vjti.ac.in (S.S. Udmale), owen.falzon@um.edu.mt (O. Falzon), g.azzopardi@rug.nl (G. Azzopardi).

https://doi.org/10.1016/j.eswa.2021.116354
Received 8 March 2021; Received in revised form 28 November 2021; Accepted 28 November 2021
Available online 12 December 2021
0957-4174/

Individual recognition of animals has also presented promising solutions in wildlife for monitoring of animals to make conservation and management decisions (Norouzzadeh et al., 2018). Various animal recognition techniques have been introduced in the literature using pattern recognition methodologies that capture the discriminatory features of animals. Notable is the fact that the visual appearance has been evaluated as an effective biometric approach to express the unique features of individual animals (Andrew, Greatwood, & Burghardt, 2017; Bhole, Falzon, Biehl, & Azzopardi, 2019; Kumar et al., 2018).

Among various animal classification approaches, cattle recognition is an active research area in animal husbandry due to its importance in livestock health monitoring, breeding practices and the food industry. The existing cattle identification systems include ear tagging, freeze branding, and the embedding of microchips approaches. Owing to their ease of use and low cost, most of these systems are operated using the labeled ear-tags with RFID devices. RFID tags function on low frequency and are able to provide vital digital information by scanning the tag electronically. As a result, the system scanner is always deployed within a few inches of the labeled ear tags for correct identification. This approach has been widely adopted around the world. It has, however, been reported that the classical animal identification approach is unable to provide reliable and effective solutions due to loss of ear-tags, fading of labels on ear-tags, and physical damage to the cattle (Andrew et al., 2017; Kumar et al., 2018). The embedded unique number has also been easily duplicated for creating false registration numbers to make fraudulent insurance claims. As a result, it cannot provide acceptable security to the owners for the verification of different breeds of animals associated with the farms and insurance firms. Conventional approaches have, however, manifested several limitations such as permanent damage (Johnston, & Edwards, 1996; Wolf, Tonsor, McKendree, Thomson, & Swanson, 2016), welfare issues (Edwards, Johnston, & Pfeiffer, 2001; Grandin, 2016), reliability (Wardrope, 1995), and high implementation cost management of large-scale monitoring of livestock animals (Feng, Fu, Wang, Xu, & Zhang, 2013). The tagging approach also fails to mitigate the demand for continuous monitoring of individual animals for addressing issues like welfare assessment, behavioral and social analysis, disease development, and infection transmission. Besides, RFID-based systems are only effective when the respective cow is very close to the scanner. Therefore, it cannot be used to monitor cattle in other areas of the farm. These limitations of conventional approaches have given rise to new research directions with various applications for animal husbandry.

Recently, the integration of new technologies with hardware, such as prioritizing sensor-based alarm systems (Dominiak, & Kristensen, 2017), has offered effective, efficient, affordable, and scalable livestock solutions for the management of essential operations. Thus, it can cater to a large number of cattle and, more importantly, monitor them using non-intrusive methods for breed association, milking, welfare concerns, control of disease outbreaks, and weighing, among other activities (Caporale, Giovannini, Di Francesco, & Calistri, 2001; Hansen, Smith, Smith, Jabbar, & Forbes, 2018; Smith, Tatum, Belk, Scanga, Grandin et al., 2005). It is also required by legal frameworks for traceability of livestock throughout their lives in order to ensure the link between the identification of animals, product development in the farm industry, and various health-related concerns (EUR-Lex, 2000; U.S. Department of Agriculture, 2020). These measures also provide protection and confidence to the consumer while addressing concerns related to public health.

There is clearly a need for more reliable and efficient automated approaches for identifying and monitoring of cattle. Biometric-based methods seem to provide the most appropriate tools to address the above challenges and concerns. However, it turns out that the generation of biometric data sets for the development of automatic methods is a very challenging task. To the best of our knowledge, there is only one publicly available data set (Andrew et al., 2017), which was collected during only one occasion of 2 h from a small herd of 89 cows. For this reason, the first step of this study was to collect a new and bigger data set from the Dairy Campus in Leeuwarden, the Netherlands, across nine occasions on which details are given in Section 3.1.

In this work a novel computer vision-based framework is proposed for the individual identification of Holstein cattle. It is based on their coat pattern as seen from the side view, and it takes as input pairs of RGB and thermal images captured simultaneously by a fixed camera. Fig. 1 illustrates a couple of examples using the Multi-Spectral Dynamic Imaging (MSX) post-processing technique developed by FLIR to informatively represent the thermal and visual spectrum's instantaneously. The advantage of MSX images is that they are developed by introducing the details of the visible spectrum to thermal images. It helps to improve the image quality along with embedding edges and lines.

The contributions of this work are three-fold. Firstly, it proposes a novel computer vision-based framework for individual identification of Holstein cattle based on their coat patterns using thermal MSX and RGB images. Secondly, it introduces a novel CORF3D contour map that accentuates the main contours of the coat pattern and suppresses texture. Thirdly, it includes a new data set of 3694 pairs of images from 383 cows and makes it available to the research community.[1] To our knowledge, it is the largest side view data set publicly available for research on Holstein cattle identification.

2. Related work

In the last decade, the field of animal biometrics has received significant attention due to the need for improving activities like monitoring, detection, recognition, and tracking of different animal species. It is worth mentioning that existing animal biometric techniques recognize animals in a non-invasive way based on the visual appearance of the animals (Kumar & Singh, 2017). Also, invasive techniques are reported in the literature to identify animals that use branding, ear-tattoo-based identification, and RFID-based ear-tagging and collar ID (Awad, 2016). Several approaches have been proposed for livestock tracking, registration, behavior analysis, and breed association (Kamilaris & Prenafeta-Boldú, 2018; Kumar & Singh, 2017; Norouzzadeh et al., 2018).

Invasive tagging-based systems are amongst the most widely used for cattle recognition but still present several limitations. A contactless and thus a non-invasive image-based biometric system can help overcome these limitations and offer an appealing approach for cattle recognition. Biometric features, such as coat patterns and muzzle prints, have already been investigated (Andrew et al., 2017; Bhole et al., 2019; Kumar et al., 2018) and considered to be very effective biometric features (Petersen, 1922). Various techniques from traditional machine learning and deep learning have been evaluated on muzzle print images to identify cattle and have shown promising results (Awad, Zawbaa, Mahmoud, Nabi, Fayed et al., 2013; Kumar et al., 2018).

Feature description algorithms such as Scale-Invariant Feature Transform (SIFT) (Lowe, 2004) have been applied on muzzle prints for cattle identification and achieved the performance of approximately 93.3% with a cattle data set of 90 muzzle prints (Awad et al., 2013). Similarly, a matching refinement technique has been proposed by utilizing the SIFT descriptor approach to recognize cattle on a data set of 160 muzzle images (Fosgate, Adesiyun, & Hird, 2006). The local texture features from such images have also been explored in combination with an Adaboost classifier (Gaber, Tharwat, Hassanien, & Snasel, 2016). The RANdom SAmple Consensus (RANSAC) approach was utilized to eliminate the outliers from the muzzle image, and they were incorporated in the SIFT algorithm for the improvement of the reliability and robustness of the proposed cattle identification approach.

[1] The data set can be downloaded from here: https://doi.org/10.34894/7M108F.

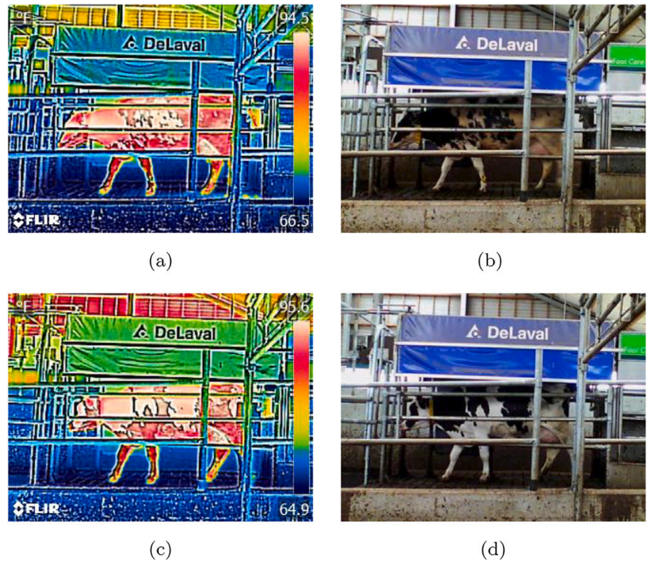


Fig. 1. Examples of pairs of thermal MSX and RGB images acquired simultaneously by the FLIR E6 thermal camera from a distance of five meters. (a–b) Images of the cow with ID 0094 and (c–d) images of the cow with ID 0099 were taken from the new data set.

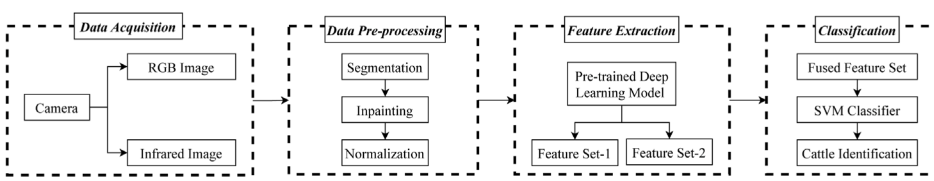


Fig. 2. Schematic overview of the proposed pipeline.

Recently, a deep learning framework has been employed for the individual identification of cattle based on muzzle print images (Kumar et al., 2018). It achieved accuracy rates up to 98.9% for tests carried out on 500 cows. While experiments show high-performance rates on good quality images, in practice, cattle may not be sufficiently cooperative with a system that requires a clear and close-up view of the muzzle.

The coat patterns of the Holstein cattle, which are exhibited as black, brown, and white patches over their bodies, have also been the subject of investigation for individual identification. In particular, an R-CNN (He, Gkioxari, Dollár, & Girshick, 2017) adaptation of the VGG-CNN-M-1024 network was proposed for this purpose, leading to a mean average precision of 86.07% over two-fold cross-validation on a data set of 89 cows using top-view still RGB images (Andrew et al., 2017). The promising results demonstrated that the coat pattern is a salient feature that can indeed be used for individual identification of Holstein cattle.

The idea of using thermal images is motivated by the benefits observed in other applications, including the recognition of proximate material type (Cho, Bianchi-Berthouze, Marquardt, & Julier, 2018), tiny face identification (Singh, Ahmed et al., 2020), and crack detection (Yang, Wang, Lin, Li, Sun et al., 2019), among others. Besides, fusion-based approaches of RGB and thermal images were also investigated in different applications for enhanced pedestrian visibility (Shopovska, Jovanov, & Philips, 2019), and traffic monitoring (Alldieck, Bahnsen, & Moeslund, 2016). In this work, the side view of the cattle is leveraged after leaving the milking parlor. The side view provides the largest surface area of a cow, and thus, it has more information as compared to the top view. Hence, this work starts by collecting a new data set with a fixed thermal camera that produces a pair of RGB and thermal images for each snapshot. The decision of using a thermal camera is mainly motivated by the facility to use the body temperature for delineating the cows from the background. Moreover, it allows the assessment of what potential temperature feature maps have with respect to the cow recognition problem at hand.

3. Methods

In this paper a novel cattle identification method is proposed. It is based on RGB and thermal data and incorporates computer vision, transfer learning, and data fusion techniques. The proposed method consists of four stages: (1) data acquisition, (2) data pre-processing, (3) feature extraction, and (4) cattle classification. The general structure of the proposed process is demonstrated in Fig. 2. Further details on each stage of the pipeline are provided below.

3.1. Data acquisition

The first step involved the collection of a new image data set with a FLIR E6 thermal camera over nine days at the Dairy Campus in Leeuwarden[2] from 383 cows. Fig. 1 shows examples of images captured for two different cows. For the majority of the cases, one image per cow per day was captured during every afternoon milking session. In the first two days of data collection, however, it became apparent that some cows may actually skip the milking session. That situation led to having less than nine images for certain cows. With this in mind, multiple images were collected for those cows that changed their posture considerably in the same session. The actual number of images (at most five) taken per cow per session depended on the time to exit the milking parlor.

Fig. 3 illustrates the distribution of samples per cow and how they were captured over the nine days.

The thermal camera uses a straightforward point and shoot mechanism. It can detect minor temperature differences around 0.06 °C and produces an image resolution of 320 × 240 pixels. Besides emissivity (0.98 in the case of cattle skin), the distance and reflected temperature

[2] https://www.dairycampus.nl/en/Home.htm

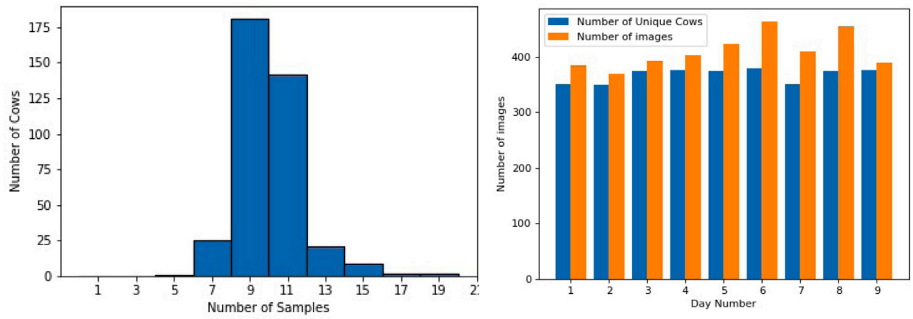


Fig. 3. (a) Distribution of samples per cow, with the majority of cows having nine images each. (b) The total number of captured images and the unique number of cows per day.

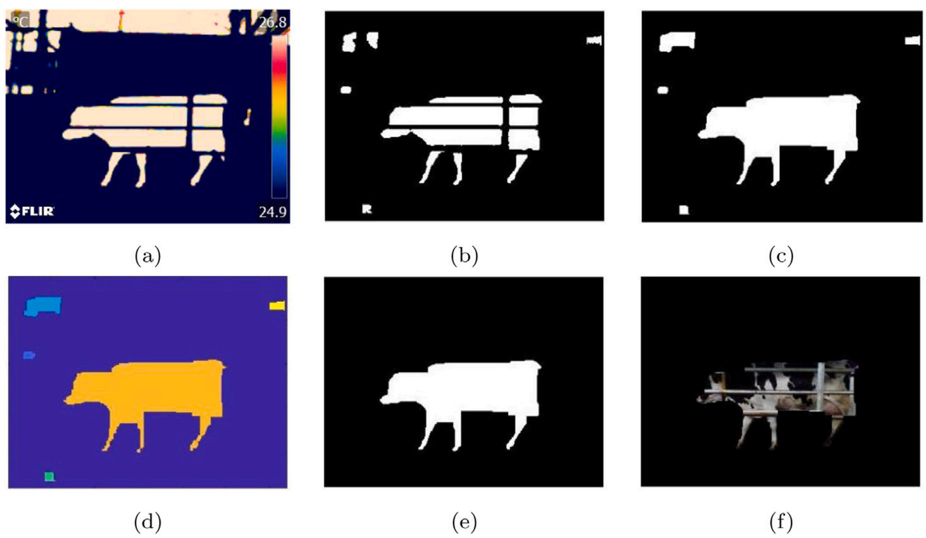


Fig. 4. Segmentation process. (a) Thermal image with clipped lower and upper values. (b) Binary image obtained after thresholding. (c) Resulting image after the morphological closing operation. (d) A connected component analysis used to extract the element with the maximum area. (e) Segmentation mask of the cattle. (f) The corresponding segmented RGB image.

were the main parameters required to be tuned during the data collection process. The camera was mounted at the height of 1.5 meters from the ground. On the first day of data acquisition, the camera was attached four meters away from the cattle. While the resulting images were of good quality, it turned out that the cows were taking almost the entire width of the pictures. Consequently, in the subsequent days, the camera was mounted at a distance of five meters from the cattle in order to ensure that the entire cow fits comfortably in the view.

3.2. Data pre-processing

The RGB and thermal images of the cattle's side views are captured just after leaving the milking barn. In order to address the challenges presented in this process, namely partial coat pattern occlusion by rods and background noise, the following three pre-processing operations were implemented; segmentation, inpainting, and normalization, which are described in detail below.

3.2.1. Segmentation

The thermal image is leveraged to segment the RGB image using the steps illustrated in Fig. 4. Initially, the thermal image is clipped between the lower and upper bound temperatures, as shown in Fig. 4(a). This range of temperatures was determined manually by visually inspecting few images.

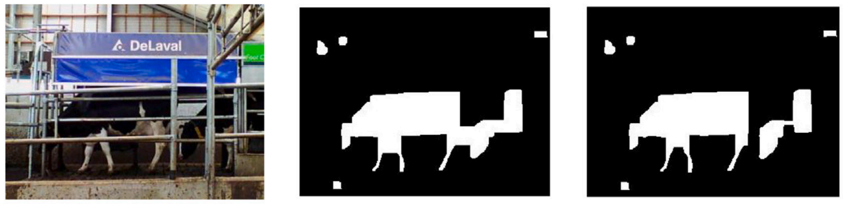
The binarized image contains surplus elements in the background and gaps across the cattle due to the presence of rods, Fig. 4(b). In order to retain the region of interest and remove pixels incorrectly labeled as foreground pixels, the closing morphological operation is applied horizontally and vertically with a line structuring element of 9 pixels, resulting in a closed binary mask as shown in Fig. 4(c). Fig. 4(d–e) illustrate the result of the connected component analysis, from which the largest element is taken to represent the cattle and the rest is ignored. The mask is then overlaid on the RGB image to crop the region of interest, Fig. 4(f).

The closing morphological operation helps to overcome one of the challenges in the data acquisition process. It concerns the occurrence of multiple cattle in the same image, where one cow manages to sneak its head out of the milking barn before the one preceding it makes way. Fig. 5 demonstrates an example of an occurrence where the cattle are in a restrictive area surrounded by rods. Thus, two cows can only be seen in line and not occluding each other. The thresholding operation explained above may result in the two cows being segmented as one large connected component.

The same segmentation procedure is applied to the thermal and MSX versions of these images.

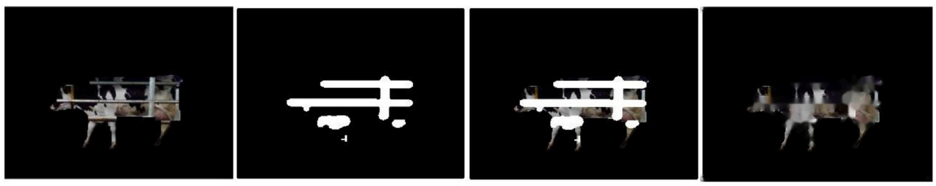
3.2.2. Inpainting

The next challenge is to address the gaps caused by the occluding rods, as shown in Fig. 6(a). This challenge is addressed by generating a binary mask for the rods, which is determined by subtracting the binary image in Fig. 4(b) from the resulting segmentation mask in Fig. 4(e). Due to the morphological closing operation that leads to the binary mask in Fig. 4(e), the resulting binary map of the subtraction operation may contain rough edges. Thereafter, a smoothing operation is applied



(a) (b) (c)

Fig. 5. Examples of multiple cattle in a single view. (a) Original RGB image and (b) the corresponding binary map with two cattle in a single component. (c) Separated cattle resulting from a closing operation with a structuring line element of 9 pixels.



(a) (b) (c) (d)

Fig. 6. (a) Segmented image of a cow. (b) Binary mask of the rods. (c) Defaced segmented RGB image. (d) Inpainted RGB image.

with a Gaussian function whose kernel size is 9×9 pixels, followed by binarization with a threshold of 0.8, as shown in Fig. 6(b).

Next, the inverted mask (Fig. 6(b)) is multiplied with the segmented RGB image in order to remove the regions covered by the rods. This process, however, defaces the segmented RGB image, as shown in Fig. 6(c). Finally, the fast marching inpainting algorithm (Telea, 2004) is applied to repair the defaced segmented RGB images using the inverted mask, resulting in the image shown in Fig. 6(d). For the inpainting algorithm, the radius of the circular neighborhood of each inpainted point is set to 3 pixels, replacing the gaps with more relevant values from the local neighboring pixels. In the same manner, the inpainting approach is applied to the thermal and MSX images using the same binary mask to fill the areas of the rods with more relevant values. Fig. 7 illustrates the sets of raw and the corresponding preprocessed images of a particular cow.

3.2.3. Normalization

The temperature maps are min–max normalized in the range [0,1], while the other feature maps are represented as images whose values are in the range [0,255].

3.3. Contour maps

The brain-inspired CORF contour detection operator (Azzopardi, & Petkov, 2012; Azzopardi, Rodríguez-Sánchez, Piater, & Petkov, 2014; Melotti, Heimbach, Rodríguez-Sánchez, Strisciuglio, & Azzopardi, 2020; Strisciuglio, Azzopardi, & Petkov, 2019) is used to extract contour maps from the RGB images. The CORF operator takes as input the responses of center-on and center-off Difference-of-Gaussians operators whose areas of support are appropriately aligned. It also includes a brain-inspired inhibition component, known as push–pull, which suppresses grain type texture and as a result accentuates the defining edges that belong to perceptually salient contours. It has two main hyperparameters, namely σ and α that represent the standard deviation of the inner Gaussian function of the afferent difference-of-Gaussians operators and the inhibition strength, respectively. The smaller the value of σ, the more finer edges are detected, and the higher the value of α, the more noise is suppressed up to a certain extent. The CORF contour operator was also extended to a keypoint detector for object localization (Azzopardi, & Petkov, 2013; Azzopardi, & Petkov, 2014; Gecer, Azzopardi, & Petkov, 2017)

A 3D CORF map is designed with each of its three channels corresponding to a different value of the inhibition strength α ($\alpha \in \{0, 1.8, 3.6\}$). Fig. 8(b–d) shows such three contour maps with the mentioned three inhibition strengths to the grayscale version of the given RGB image in Fig. 8(a).

3.4. Feature fusion and the classification model

Due to a relatively small data set, which consists of an average of nine images per cow, a transfer learning approach is preferred to determine a classification model. A pre-trained convolution neural network (ConvNet) model is initialized with its last layer replaced from the 1000 ImageNet classes to the 383 cow classes at hand.

All the layers of the given pre-trained ConvNet are unfrozen and fine-tuned with training images from the new Holstein cattle data set. In this work, four different feature sets are considered, which are extracted from RGB, MSX, temperature and contour image maps. Moreover, two types of classification approaches are investigated. In the first approach, the four feature sets are used as input, and the output of the respective ConvNet determines the label of the given cow. In the second approach, feature fusion is used by concatenating the vectors in the fully connected (FC) layers of two ConvNets with the same architecture, but that take different feature sets as input, as shown in Fig. 9. Subsequently, the concatenated vector is fed to a linear SVM.

4. Experiments

This section describes the experiments and provides a comparative analysis of several ConvNets and the fusion of different feature sets.

4.1. Data set

Four different feature sets are constructed from the RGB, thermal, MSX, and contour maps with three channels each. The red, green, and blue color maps are the three channels representing the RGB and MSX images. The temperature map is triplicated, and the resulting 3D map is denoted by Temp3D. Moreover, three contour maps, denoted by CORF3D, are produced with three different values of the inhibition strength parameter α. The data set consists of a total of 3694 images from 383 classes (i.e., cows) with an average of 9.6 images per category. Fig. 10 illustrates the representative RGB, thermal MSX, temperature, and contour maps of one example cow from the data set.

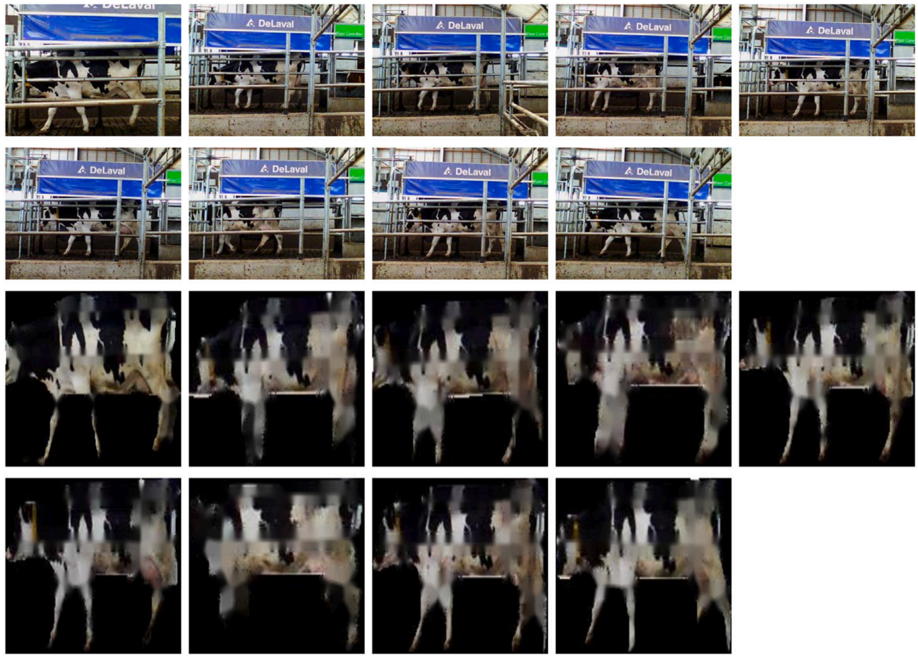
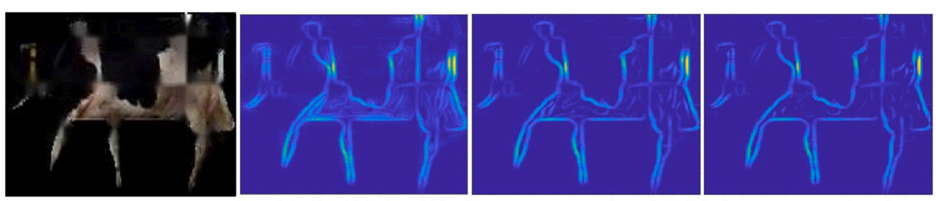


Fig. 7. (Top two rows) The original RGB and (bottom two rows) the corresponding preprocessed images of cow 0099. All images were taken on different days over the nine-day data acquisition period. The first RGB image, which was acquired on day 1, was captured from a distance of 4 meters and the remaining images from a distance of 5 m. Nevertheless, the preprocessing step crops the respective cattle and resizes them to the same dimensions.



(a) (b) (c) (d)

Fig. 8. (a) Segmented and inpainted cattle in RGB and the corresponding CORF contour maps with standard deviation σ = 2.2 and different values of the parameter α that determines the inhibition strength: (b) $\alpha = 0$ (c) $\alpha = 1.8$ (d) $\alpha = 3.6$.

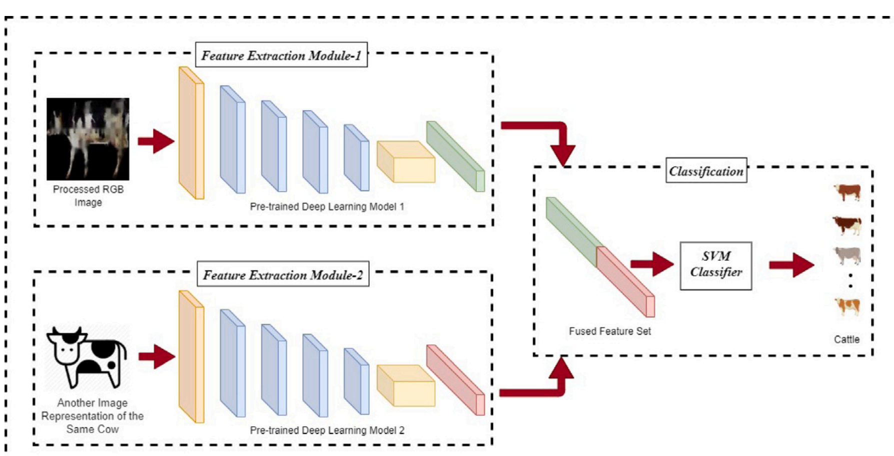
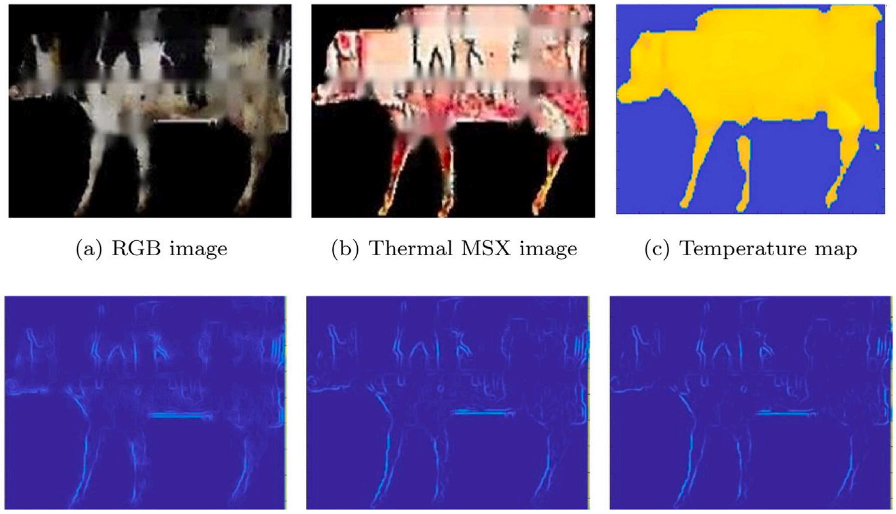


Fig. 9. Schematic overview of the fusion approach. The architectures of the pre-trained deep learning models are the same.

4.2. Experimental setup

Three ConvNets, namely MobileNet (Howard et al., 2017), Xception (Chollet, 2017), and DenseNet121 (Huang, Liu, Van Der Maaten, & Weinberger, 2017) are evaluated, which are pre-trained on the ImageNet data set (Deng, Dong, Socher, Li, Li et al., 2009), and are fine tuned individually with each of the four feature sets mentioned above. These specific three modeling architectures are chosen based on the criteria to provide a fast, efficient and effective solution for the system. The following hyperparameters are used to fine-tune the



(a) RGB image (b) Thermal MSX image (c) Temperature map

(d) CORF map: $\alpha = 0$ (e) CORF map: $\alpha = 1.8$ (f) CORF map: $\alpha = 3.6$

Fig. 10. Examples of RGB, thermal MSX, temperature map and CORF maps for the cattle with ID = 0099.

pre-trained models: learning rate = 0.0005, decay = 0.00001, categorical cross-entropy loss function, Adam optimizer, and early stopping with 15 epochs. The implementation is based on standard Python libraries[3].

A 5-fold cross-validation approach is used for each experiment, and the average results with the corresponding standard deviations are reported. First, the proposed approach is evaluated using the four feature sets individually. It is followed by investigating the effect of fusing specific pairs of feature sets.

In order to further evaluate the generalization ability of the proposed approach another experiment is considered. It is referred to as leave-one day-out cross-validation, and involves nine runs (one for each day of the data collection), with each run having a test set that contains all images of all cows collected on a specific day. For instance, run 1 uses all images collected on the first day for testing and all images of the remaining eight days for training, run 2 uses all images captured on the second day for testing, and so forth.

4.3. Results

Table 1 shows the results achieved when the feature sets are used individually using the 5-fold cross-validation approach. It is observed that all three types of ConvNet models achieve an average accuracy higher than 98.0% with the RGB-based features. In particular, DenseNet121 attains marginally better performance than its counterparts, also with a low standard deviation. This demonstrates the stability of the DenseNet121 model for the RGB-based feature set. Similarly, for the MSX-based feature sets, the accuracy of the model is higher than 95.0% but approximately 3.0% lower than that achieved with the RGB-based features. The best performance for the MSX-based features is achieved by MobileNet, which obtains an accuracy of 97.04%. These results so far suggest that the RGB-based feature sets provide distinguishing patterns for the classification of cattle. The Temp3D features extracted from the temperature maps did not yield satisfactory results by any of the models.

Notable is the fact that the overall best accuracy of 99.16% (also with the smallest standard deviation) is achieved with the CORF contour maps extracted from the grayscale versions of the RGB images combined with the MobileNet model. A high accuracy rate with the same feature sets is also achieved with the Xception model.

Table 1
Average performance of three ConvNets and four feature sets across 5-fold cross-validation.

Model	Feature set	Test accuracy (%)	F1 (%)
MobileNet	RGB	98.18 ± 0.28	97.75 ± 0.25
	MSX	97.04 ± 0.37	96.88 ± 0.42
	Temp3D	50.71 ± 1.34	50.15 ± 1.24
	CORF3D	99.16 ± 0.21	99.08 ± 0.18
DenseNet121	RGB	98.56 ± 0.28	98.51 ± 0.30
	MSX	95.38 ± 0.69	95.34 ± 0.55
	Temp3D	13.12 ± 4.23	6.95 ± 4.56
	CORF3D	97.88 ± 0.27	98.27 ± 0.25
Xception	RGB	98.36 ± 0.28	98.34 ± 0.24
	MSX	98.36 ± 0.28	98.37 ± 0.36
	Temp3D	9.38 ± 2.34	6.01 ± 3.68
	CORF3D	98.86 ± 0.31	98.75 ± 0.35

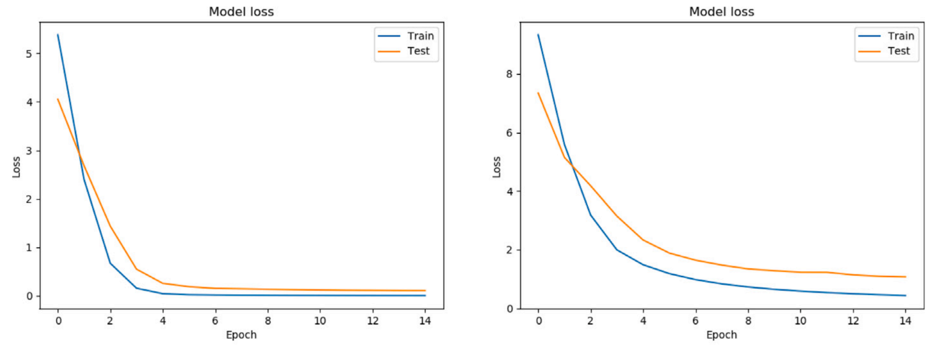
Table 2
Average performance of different fusions of feature sets across 5-fold cross-validation.

Model	Feature set	Test accuracy (%)	F1 (%)
MobileNet + SVM	RGB + MSX	98.68 ± 0.35	98.48 ± 0.38
	RGB + CORF3D	99.19 ± 0.11	99.11 ± 0.15
DenseNet121 + SVM	RGB + MSX	99.00 ± 0.41	98.96 ± 0.48
	RGB + CORF3D	**99.64 ± 0.13**	**99.54 ± 0.16**
Xception + SVM	RGB + MSX	98.1 ± 0.14	98.08 ± 0.18
	RGB + CORF3D	99.29 ± 0.37	99.24 ± 0.33

Table 2 reports the performance of different pairs of fused feature sets. The RGB features are used in all fused feature sets due to their stable performance of approximately 98.0% for all pre-trained models. Besides, the temperature maps are not considered for further analysis due to its unsatisfactory performance, as reported in Table 1.

The results show that the feature fusion approach reduces the error rate in all cases except for the set of RGB and MSX features combined with the Xception model. All three models achieve an accuracy of more than 99.0% with the fusion of features extracted from the RGB and the corresponding CORF contour maps. The best overall average accuracy is 99.64% and it is achieved with the fusion of RGB and CORF3D features combined with the DenseNet121 model. The small SD value of 0.13 also indicates the stability of the proposed model.

[3] The proposed pre-trained deep learning framework is implemented using the Keras library, and the scikit-learn library is utilized for the SVM classifier. The experiments were executed on a Nvidia V100 GPU card.



(a) RGB features (b) CORF3D features

Fig. 11. The loss functions of the DenseNet121 model for the training and test data for the first 14 epochs using (a) RGB and (b) CORF3D features separately.

Table 3
Average performance of different feature sets and models across a leave-one day-out cross-validation.

Model	Feature set	Test accuracy (%)	F1 (%)
MobileNet	RGB	98.01 ± 0.39	98.08 ± 0.45
MobileNet	CORF3D	98.43 ± 0.47	99.55 ± 0.26
MobileNet + SVM	RGB + CORF3D	99.64 ± 0.40	99.44 ± 0.42
DenseNet121	RGB	98.86 ± 0.45	98.75 ± 0.43
DenseNet121	CORF3D	98.85 ± 0.39	98.76 ± 0.50
DenseNet121 + SVM	**RGB + CORF3D**	**99.71 ± 0.30**	**99.68 ± 0.30**
Xception	RGB	98.46 ± 0.64	98.66 ± 0.28
Xception	CORF3D	98.60 ± 0.53	98.59 ± 0.26
Xception + SVM	RGB + CORF3D	99.15 ± 0.35	99.21 ± 0.33

The learning quality of the classifiers is illustrated in Fig. 11, which shows the loss functions using the RGB and CORF3D features for DenseNet121 as achieved by one of the five folds. These plots demonstrate the convergence of the models with no signs of under- or over-fitting.

Table 3 reports the results of the leave-one day-out cross-validation experiment. The results are very consistent with those achieved by the 5-fold cross-validation reported in Tables 1 and 2. In fact, the best performance is achieved with the same model and the same combined feature set of RGB and CORF3D. These results, therefore, confirm that the proposed model has a very strong generalization ability.

5. Discussion

To the best of our knowledge, this is the first study that proposes a solution for cattle recognition from the side view and that has been evaluated on a new and reasonably large data set of 3694 images and 383 cows. The experiments tested four different feature sets and investigated their individual and combined effectiveness for the identification of cattle. The best accuracy rates of 99.64% and 99.71% are achieved with the fusion of RGB and CORF3D features for the five-fold and leave-one day-out cross-validation experiments, respectively. The additional CORF3D features reduced the error rate by 75.0% in both cases when compared to the respective accuracy rates of 98.56% and 98.86% achieved by the RGB features only. This significant improvement demonstrates the quality of the CORF3D features that concern the contours of the coat patterns with different levels of noise suppression. It can be speculated that this approach of combining RGB features with CORF3D features extracted from the same image can be effective in a broad range of computer vision applications.

Fig. 12 illustrates all ten misclassified samples over the nine runs of the leave-one day-out cross-validation. The majority of the misclassified cases involve cattle which are almost completely black, brown or white, and thus have few or no patterns that characterize them. Of the ten misclassified samples, two belong to the same cow 0444 and they are mistakenly classified with the same wrong class (cow 7901). Conversely, there is one sample from cow 7901, which has been wrongly classified with cow 0444. These misclassifications occurred due to the lack of defining piebald patterns in the respective cows. In order to address the challenge that such cows present, one may consider a multi-camera system that can capture the side, top, frontal and rear views and investigate a modeling approach that can take all views as input. Such a system would be able to leverage more distinctive patterns from as many body parts as possible. Another challenge that may occur in practice is considerable dirt that may cover some of the defining piebald features. The new data set of 383 cows that was collected for this study does not contain samples with severe dirt, and hence the proposed method could not be evaluated in such situations. This is where the multi-view system mentioned above could also play an important role in addressing this issue. Relying on multiple views would potentially decrease the effect that a dirty side of a cow may have on the recognition method.

The proposed framework can be installed in farms for cattle recognition from the side view in milking stations. It is not dependent on a specific thermal camera. The thermal image is only used to segment the cow from the background. Subsequently, the standard RGB images and the respective CORF contour maps are used as input to the model. Due to the way the data set was collected, the proposed solution includes a preprocessing step that detects the occluding rods and replaces them with an inpainting algorithm. While this step introduces few artifacts, it turned out to be more effective than keeping the occluding rods in view. In fact, when skipping this step, the proposed fusion methodology that uses both RGB and CORF3D feature maps coupled with DenseNet121 and SVM leads to an accuracy of 99.11% (± 0.53). The current framework can be extended to provide integrated solutions for other types of animals who have distinctive visual features. It may also provide the opportunity to device manufacturers to design their platforms and sensors to achieve the task. As a result, it will contribute to transforming the agricultural industry by modernizing livestock management systems.

Due to the limited number of samples available per cow, this work uses transfer learning and fine-tunes the networks according to the number of classes (i.e., cattle) in the new data set. In the future, one may develop a siamese network to learn a similarity function between samples that belong to the same cows versus the rest. Such an approach would not require retraining the models every time there are changes in the cattle of a given farm.

One limitation of the proposed method is relying on a single snapshot for the identification of a cow. Due to the occluding rods at the exit of the milking station, there can be a situation where the occluded pattern is the most salient feature that distinguishes that cow from another. To address this issue, future work may consider taking multiple snapshots of the same cow while exiting the milking station. This approach would give the opportunity for all salient features to be visible in at least one of the images. By applying a suitable fusion approach, the effect of occluding rods would be reduced further.

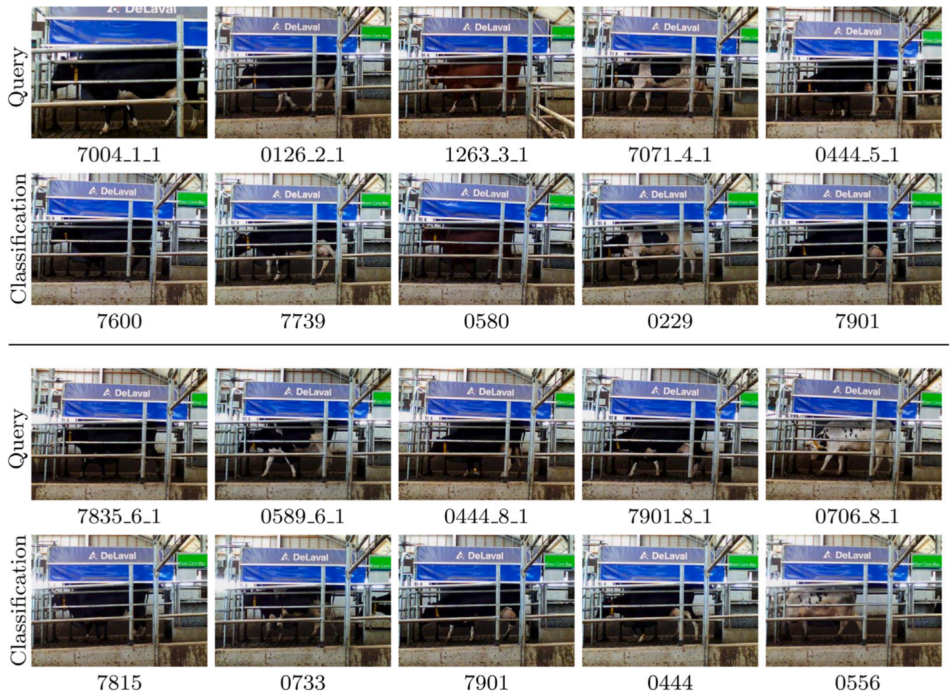


Fig. 12. All ten misclassified samples in the leave-one day-out cross-validation over the nine folds. The first and second rows of each of the two sections show the query images that were misclassified, and example images from the misidentified classes, respectively. The labels under the query images have the following format: <cow ID> _ <day no.> _ <sample no.>. The other labels indicate the cow IDs of the misidentified cows.

6. Conclusion

This study demonstrates that the automatic identification of individual Holstein cattle from the side view can be performed using a computer vision and machine/deep learning pipeline, consisting of image acquisition, segmentation, inpainting, feature extraction, and transfer learning. By means of experiments on a new large data set, it can be concluded that the proposed approach is very effective for cattle identification. The best accuracy rates of 99.64% and 99.71% are achieved by late fusion of DenseNet121 FC features fine-tuned on RGB and CORF3D contour maps, using 5-fold cross-validation and leave-one day-out approaches, respectively. These results are obtained on the new 383-class data set with 3694 images. The novel CORF3D contour maps have proven to give the most substantial contribution to the results.

The proposed approach is entirely non-invasive and non-intrusive as it relies on visual information only. This also means that the cattle would not need to carry sensors on them, nor would they require to pose in specific postures. In a practical setting, this method would only require one fixed camera connected to a central server or to the cloud where all the processing takes place. The proposed non-invasive and non-intrusive approach can, therefore, serve to stimulate further research for the identification of cattle anywhere and anytime on a given farm. This will further improve the well-being of such animals while allowing the ability of enhanced phenotyping.

CRediT authorship contribution statement

Amey Bhole: Data curation, Methodology, Validation, Software, Investigation, Resources, Writing – original draft, Writing – review & editing. **Sandeep S. Udmale:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Validation, Investigation. **Owen Falzon:** Conceptualization, Methodology, Resources, Writing – original draft, Writing – review & editing. **George Azzopardi:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Supervision, Project administration.

Declaration of competing interest

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

Acknowledgment

We thank Kees de Koning and his team at the Dairy Campus in Leeuwarden for permitting the data collection used in this project and for approving its availability for academic use. We are also grateful for the discussions held with Michael Biehl from the University of Groningen who contributed in the designing stage of several aspects of the project. Finally, we thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Peregrine high performance computing cluster.

References

Alldieck, T., Bahnsen, C. H., & Moeslund, T. B. (2016). Context-aware fusion of RGB and thermal imagery for traffic monitoring. *Sensors*, *16*(11), 1947.

Alsahaf, A., Azzopardi, G., Ducro, B., Hanenberg, E., Veerkamp, R. F., & Petkov, N. (2018). Prediction of slaughter age in pigs and assessment of the predictive value of phenotypic and genetic information using random forest. *Journal of Animal Science*, *96*(12), 4935–4943.

Alsahaf, A., Azzopardi, G., Ducro, B., Hanenberg, E., Veerkamp, R. F., & Petkov, N. (2019). Estimation of muscle scores of live pigs using a kinect camera. *IEEE Access*, *7*, 52238–52245.

Andrew, W., Greatwood, C., & Burghardt, T. (2017). Visual localisation and individual identification of holstein friesian cattle via deep learning. In *Proceedings of the IEEE international conference on computer vision workshops* (pp. 2850–2859).

Awad, A. I. (2016). From classical methods to animal biometrics: A review on cattle identification and tracking. *Computers and Electronics in Agriculture*, *123*, 423–435.

Awad, A. I., Zawbaa, H. M., Mahmoud, H. A., Nabi, E. H. H. A., Fayed, R. H., & Hassanien, A. E. (2013). A robust cattle identification scheme using muzzle print images. In *2013 Federated conference on computer science and information systems* (pp. 529–534). IEEE.

Azzopardi, G., & Petkov, N. (2012). A CORF computational model of a simple cell that relies on LGN input outperforms the gabor function model. *Biological Cybernetics*, *106*(3), 177–189.

Azzopardi, G., & Petkov, N. (2013). Trainable COSFIRE filters for keypoint detection and pattern recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(2), 490–503. http://dx.doi.org/10.1109/TPAMI.2012.106.

Azzopardi, G., & Petkov, N. (2014). Ventral-stream-like shape representation: from pixel intensity values to trainable object-selective COSFIRE models. *Frontiers in Computational Neuroscience*, *8*, 80. http://dx.doi.org/10.3389/fncom.2014.00080, URL: https://www.frontiersin.org/article/10.3389/fncom.2014.00080.

Azzopardi, G., Rodríguez-Sánchez, A., Piater, J., & Petkov, N. (2014). A push-pull CORF model of a simple cell with antiphase inhibition improves SNR and contour detection. *PLoS One*, *9*(7), Article e98424.

Bhole, A., Falzon, O., Biehl, M., & Azzopardi, G. (2019). A computer vision pipeline that uses thermal and RGB images for the recognition of holstein cattle. In *International conference on computer analysis of images and patterns* (pp. 108–119). Springer.

Caporale, V., Giovannini, A., Di Francesco, C., & Calistri, P. (2001). Importance of the traceability of animals and animal products in epidemiology. *Revue Scientifique Et Technique-Office International Des Epizooties*, *20*(2), 372–378.

Cho, Y., Bianchi-Berthouze, N., Marquardt, N., & Julier, S. J. (2018). Deep thermal imaging: Proximate material type recognition in the wild through deep learning of spatial surface temperature patterns. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1–13).

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1251–1258).

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *CVPR09*.

Dominiak, K. N., & Kristensen, A. R. (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*, 46–67.

Edwards, D., Johnston, A., & Pfeiffer, D. (2001). A comparison of commonly used ear tags on the ear damage of sheep. *Animal Welfare*, *10*(2), 141–151.

EUR-Lex (2000). European union law. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32000R1760 (Accessed 03 February 2021).

Feng, J., Fu, Z., Wang, Z., Xu, M., & Zhang, X. (2013). Development and evaluation on a RFID-based traceability system for cattle/beef quality safety in China. *Food Control*, *31*(2), 314–325.

Fosgate, G., Adesiyun, A., & Hird, D. (2006). Ear-tag retention and identification methods for extensively managed water buffalo (*Bubalus bubalis*) in Trinidad. *Preventive veterinary medicine*, *73*, 287–296. http://dx.doi.org/10.1016/j.prevetmed.2005.09.006.

Gaber, T., Tharwat, A., Hassanien, A. E., & Snasel, V. (2016). Biometric cattle identification approach based on weber's local descriptor and adaboost classifier. *Computers and Electronics in Agriculture*, *122*, 55–66.

Gecer, B., Azzopardi, G., & Petkov, N. (2017). Color-blob-based COSFIRE filters for object recognition. *Image and Vision Computing*, *57*, 165–174. http://dx.doi.org/10.1016/j.imavis.2016.10.006, URL: https://www.sciencedirect.com/science/article/pii/S0262885616301895.

Grandin, T. (2016). Evaluation of the welfare of cattle housed in outdoor feedlot pens. *Veterinary and Animal Science*, *1*, 23–28.

Hansen, M. F., Smith, M. L., Smith, L. N., Jabbar, K. A., & Forbes, D. (2018). Automated monitoring of dairy cow body condition, mobility and weight using a single 3D video capture device. *Computers in Industry*, *98*, 14–22.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2961–2969).

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4700–4708).

Johnston, A., & Edwards, D. (1996). Welfare implications of identification of cattle by ear tags. *Veterinary Record*, *138*(25), 612–614.

Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, *147*, 70–90.

Kumar, S., Pandey, A., Satwik, K. S. R., Kumar, S., Singh, S. K., Singh, A. K., & Mohan, A. (2018). Deep learning framework for recognition of cattle using muzzle point image pattern. *Measurement*, *116*, 1–17.

Kumar, S., & Singh, S. K. (2017). Visual animal biometrics: survey. *IET Biometrics*, *6*(3), 139–156.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2), 91–110.

Melotti, D., Heimbach, K., Rodríguez-Sánchez, A., Strisciuglio, N., & Azzopardi, G. (2020). A robust contour detection operator with combined push-pull inhibition and surround suppression. *Information Sciences*, *524*, 229–240. http://dx.doi.org/10.1016/j.ins.2020.03.026, URL: https://www.sciencedirect.com/science/article/pii/S0020025520302073.

Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., & Clune, J. (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences*, *115*(25), E5716–E5725.

Petersen, W. (1922). The identification of the bovine by means of nose-prints. *Journal of Dairy Science*, *5*(3), 249.

Shopovska, I., Jovanov, L., & Philips, W. (2019). Deep visible and thermal image fusion for enhanced pedestrian visibility. *Sensors*, *19*(17), 3727.

Singh, R., Ahmed, T., Singh, R., Udmale, S. S., & Singh, S. K. (2020). Identifying tiny faces in thermal images using transfer learning. *Journal of Ambient Intelligence and Humanized Computing*, *11*(5), 1957–1966.

Singh, S. K., Singh, R. S., Pandey, A. K., Udmale, S. S., & Chaudhary, A. (2020). *IoT-based data analytics for the healthcare industry: Techniques and applications*. Academic Press.

Smith, G., Tatum, J., Belk, K., Scanga, J., Grandin, T., & Sofos, J. (2005). Traceability from a US perspective. *Meat Science*, *71*(1), 174–193.

Strisciuglio, N., Azzopardi, G., & Petkov, N. (2019). Robust inhibition-augmented operator for delineation of curvilinear structures. *IEEE Transactions on Image Processing*, *28*(12), 5852–5866. http://dx.doi.org/10.1109/TIP.2019.2922096.

Telea, A. (2004). An image inpainting technique based on the fast marching method. *Journal of Graphics Tools*, *9*(1), 23–34.

U. S. Department of Agriculture (2020). Animal and plant health inspection service, U.S. department of agriculture. https://www.aphis.usda.gov/aphis/ourfocus/animalhealth/nvap/NVAP-Reference-Guide/Animal-Identification/Cattle-Identification (Accessed 03 February 2021).

Wardrope, D. (1995). Problems with the use of ear tags in cattle. *The Veterinary Record*, *137*(26), 675.

Wolf, C., Tonsor, G., McKendree, M., Thomson, D., & Swanson, J. (2016). Public and farmer perceptions of dairy cattle welfare in the united states. *Journal of Dairy Science*, *99*(7), 5892–5903.

Yang, J., Wang, W., Lin, G., Li, Q., Sun, Y., & Sun, Y. (2019). Infrared thermal imaging-based crack detection using deep learning. *IEEE Access*, *7*, 182060–182077.