Outdoor animal tracking combining neural network and time-lapse cameras

Mathieu Bonneau, Jehan-Antoine Vayssade, Willy Troupé, Rémy Arquet

HAL Id: hal-02624666
https://hal.inrae.fr/hal-02624666

Submitted on 26 May 2020

Outdoor animal tracking combining neural network and time-lapse cameras

Mathieu Bonneau[a],[*], Jehan-Antoine Vayssade[a], Willy Troupe[b], Rémy Arquet[b]

[a] URZ Zootechnique Research, INRA, 97170 Petit-Bourg (Guadeloupe), France
[b] UEPTEA Tropical Platform for Animal Experimentation, INRA, 97170 Petit-Bourg (Guadeloupe), France

ABSTRACT

Monitoring animal location can be a valuable tool in research and for practical applications, such as health or pasture management. Although GPS is commonly used, other solutions are available, such as RFID or image analysis. Image analysis is a non-invasive technique that has been proved to be useful to monitor animal location, as well as animal behaviour. Most, if not all, applications of image analysis for the continuous monitoring of farm animals have been developed with top-view cameras in indoor conditions. In this article, we develop a framework that combines low cost time lapse cameras, machine learning, and image registration, in order to monitor the location of animals in a pasture. We tested our framework by monitoring two flocks of goats under farm-like conditions. One time lapse camera was able to monitor an area of approximately 20 m by 20 m, and several cameras were combined to monitor the entire pasture. The precision and sensitivity of this method for automatic animal detection was estimated to be 90% and 84.5%, but the results can vary with the layout of the pasture. For example, goats were hardly detectable in front of a natural hedge, which appears dark in the image. In addition, any unwanted elements in the pasture can increase the false positive detection rate. Small animals, such as kids, were also difficult to detect in some cases, as they can be smaller than the weeds. With all the tested layouts, the sensitivity varies from 70.7% to 94.8% and the precision varies from 83.8% to 95.6%. The spatial accuracy of the method was also estimated. At a distance of 10 m, the maximal accuracy is approximately 56 cm, whereas the maximal accuracy is equal to 116 cm when the animals are at a distance of 20 m from the camera. This study shows that image analysis can be an interesting alternative to GPS with comparable accuracy and significantly lower cost.

1. Introduction

Affordable sensors are now widely available and offer great opportunities to farmers and researchers. For animal management, sensors could be the basis of efficient decision tools at the individual scale. One of the most exciting perspectives is to be able to monitor animal health in real time and eventually allow the early detection of disease, which could reduce the spread of the disease and limit the use of chemical treatments (González et al., 2008). From a research perspective, sensors offer the possibility of collecting a large amount of quantitative data in terms of animal physiology, health, or behaviour (Nasirahmadi et al., 2017). In this article, we will be interested in monitoring animals to record their positions over time.

To record an animal's position, several tools are available, such as RFID (Adrion et al., 2018), GPS (Turner et al., 2000), acoustic tags (Kolarevic et al., 2016), and computer vision (Brünger et al., 2018). With RFID, the animals are set up with a transponder, and antennas are

distributed over the study area. When the animal is in a particular area, its transponder is detected by one of the antennas and the animal is known to be at a certain distance from the antenna. In practice, this affords poor spatial accuracy and it is generally better suited to monitoring the visits of animals to a particular zone, such as a watering hole or feeding point (Adrion et al., 2018). To monitor larger areas, GPS is obviously one of the most common choices, although the sampling frequency and spatial accuracy can be highly variable (Buerkert and Schlecht, 2009; Swain et al., 2011). Enhancements of GPS include differential GPS, which uses ground-based reference stations in addition to satellites to increase the spatial accuracy. Acoustic tags are mostly used in fisheries and also rely on triangulation between the tags and the receivers. For all these methods, the animals must be set up with a tag, which is not always desirable, as it can be time consuming to manage and set up the tags, as well as painful for the animal. More generally, the labour and cost increase with the number of studied animals. In contrast, computer vision offers a non-invasive and cost-effective
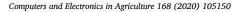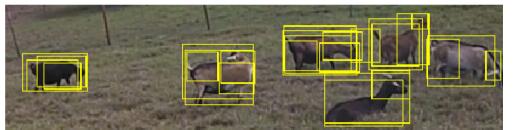
(a)

(b)

method to record the positions of animals.

Monitoring using computer vision means that images are captured and then automatically analysed to extract the desired information. Images can be captured via a simple RGB camera (Kashiha et al., 2014), a CCVT camera (Nasirahmadi et al., 2017), an infrared camera (Zhou et al., 2017), a 3D camera (Kongsro, 2014; Mortensen et al., 2016), or a depth camera (Leonard et al., 2019).

There are mainly two types of methods to detect animals in a picture. The first one consists of combining image enhancement techniques (Mahajan and Dogra, 2015) with segmentation techniques (Pal and Pal, 1993), to move from the original image to a black and white image where the animals and the background are differentiated. Then, so that the body of each animal is localized, an ellipse-fitting algorithm is generally applied. The parameters of the ellipse can then be used to derive interesting information. For example, the centroids of the ellipses can be used to determine the positions of the animals, and the lengths of the major and minor axes and the orientations of the ellipses can be used to detect individual activity, locomotion, or behaviour (Kashiha et al., 2014; Nasirahmadi et al., 2017; Vayssade et al., 2019). The second method consists of using a neural network that has been trained to identify and label the objects in the image. Neural networks, and more generally deep learning, have been used in numerous vegetal applications (Kamilaris and Prenafeta-Boldú, 2018), and animal applications (Kashiha et al., 2014; Fukunaga et al., 2015; Villa et al., 2017).

Most of the studies on animal monitoring using computer vision have been conducted for pigs, in inside conditions with top-view cameras. Although dirt and insects can obstruct the camera, this set-up is generally convenient, as it limits occlusions between animals, and also because the environmental and lighting conditions do not change drastically. Electricity and networks are also generally available to record the images at a high frequency (e.g. 25 fps) for several hours or days. Studies on monitoring an entire flock in a pasture using computer vision in a continuous or semi-continuous way have not yet been reported, although Benvenutti et al. (2015) showed that images can be used to manually count and locate cattle at a watering point. In contrast, monitoring from remotely sensed imagery (Hollings et al., 2018) has been used several times. But it is more suited for monitoring animal populations at a large scale, as, for example, counting the population of an endangered species (Fretwell et al., 2014). At the farm scale, monitoring from ground cameras appears to be more practical.

To be able to record pictures of the flock under outside conditions, we propose using time-lapse construction cameras that can run for

several weeks in outside conditions with classical batteries. Since the background is constantly changing, classical segmentation techniques are hard to apply, and we chose using a state of the art convolutional neural network, named Yolo (Redmon et al., 2016). We used Yolo to detect any object in the images, not only the animals. Then all the detected objects are filtered in order to keep only the studied animals. Finally, we set up a framework to be able to estimate the positions of the animals, in meters, in the pasture, from the positions of the animals' centroids in the image, in pixels.

2. Materials & methods

2.1. Animal monitoring framework

2.1.1. Image acquisition

We used construction time-lapse cameras (TLC2000 pro, year 2018, brand Brinno) equipped with waterproof plastic protection. These cameras record at 1.3 Mpx with a resolution of 1208×720 using jpeg compression. The cameras run on four classical AA batteries and can handle a 32 Go SD card that can store up to 240,000 pictures. The camera can be set up with different picture frequencies, from one second to 24 h. The main advantage of these cameras is their battery life, which reduces the need for human intervention. Examples from the user manual report that during daylight, with one picture every 30 s, the batteries are supposed to last for 29 days, and 17 days for a frequency of 1 image every 10 s.

2.1.2. Object detection

For object detection, we used a pre-trained convolutional neural network named Yolo (Redmon et al., 2016). More precisely, we used the network tinyYoloV3, which had been trained on the Pascal VOC data-set. We choose Yolo for its rapidity, but other available networks with higher precision can be used. For each image, Yolo was run to detect every object in the image. The output is then a set of bounding boxes (bboxes) around the detected objects. Multiple bboxes can be associated to the same object. For example, one can be around the head of the animal while another can be associated to other body parts. To merge the overlapping bboxes for the same object, we used a common non-max suppression technique. An example is presented in Fig. 1.

With a NVidia Tesla C2075 gpu, the animal detection took on average 0.28 s per image.

2.1.3. Object classification

Yolo is trained to classify common objects such as bikes, dogs, etc., but was not specifically trained to classify farm animals. As a result, the Yolo output contains false positive detections, such as birds, fences or any objects that can be found in the pasture. We thus faced an image classification problem, with the two classes *goat* and *other*. In order to be able to distinguish between goats and non-goats, we used bag of visual words to extract image features from the two classes, and then an SVM classifier. Note that any type of feature and classifier can be used to filter the detections made by Yolo. One can also use the transfer learning method to train any available CNN for image classification.

2.1.4. Locations of animals in the pasture

At this stage, all the animals are theoretically detected: (i) objects are detected by Yolo and (ii) are then filtered with a category classifier to keep only the studied animals. Each detected animal is associated to a bbox with pixel coordinates of the top left corner (px, py), width w and height h.

We defined the animal's location as the projection of the centroid on the pasture ground. Depending on the animal's orientation, it is hard to define mathematically the pixel coordinates of the projection, and we proposed using the following approximation:

$$ppx = px + \frac{w}{2} \text{ and } ppy = py + 0.9 * h, \tag{1}$$

where (ppx, ppy) are the pixel coordinates of the projection. Note that in our case, the top left corner of the image has pixel coordinates $(0, 0)$ and the bottom right corner $(1280, 720)$. See Fig. 1(b) for an example.

When monitoring animals in outdoor conditions, it is much easier to use side-view cameras than top-view ones. In this case, the distances between the objects in the image appear different from reality, due, for example, to lens distortion and perspective effects. To be able to locate the animal in the field, it is necessary to use a geometric transformation f, also called a mapping function, such that $(x, y) = f(ppx, ppy)$ would be the spatial coordinates of the animal in the pasture. This can be framed as an image registration problem, where a moving image has to be aligned with a fixed image that has known spatial coordinates. In our case, we distributed n_m marks on the pasture ground with known spatial coordinates, denoted by $\{(mx_i, my_i)\}_{i=1, ..., n_m}$. Then each mark is identified from one image from a time-lapse camera and the mark's pixel coordinates, denoted by $\{(mpx_i, mpy_i)\}_{i=1, ..., n_m}$, are recorded as follows:

$$(mx_i, my_i) = f(mpx_i, mpy_i), \text{ for } i = 1, ..., n_m. \tag{2}$$

Now, one possibility is to fix a given form for the mapping function f and estimate its parameters using the set of equations defined in (2). For the type of the mapping function f, we used a projective transformation, which provides satisfactory results, but other choices are possible (Goshtasby, 1988; Zitova and Flusser, 2003).

The value of f depends on the type of camera as well as its position. In practice, we always used the same cameras at the same positions. Thus, the marks' pixel coordinates and f have to be computed only once, at the beginning of the monitoring period, and f can be used to determine the animals' spatial coordinates based on the positions of their images (ppx, ppy).

A schematic representation of the entire workflow for the animal monitoring is presented in Fig. 2.

2.2. Experimental evaluation

2.2.1. Animal detection

We tested our monitoring framework in several situations. All the experiments were conducted at the INRA-PTEA farm located in Guadeloupe.

2.2.1.1. Long term monitoring of nursing goats D_1.
We selected a flock of approximately 18 goats and 22 kids, managed under rotational grazing.

The flock is grazing across 5 different pastures, spending one week per pasture. We set up the cameras and marks on one pasture from February 2019 to April 2019 and we thus monitored the flock for three weeks in all. The pasture was approximately 1300 square meters and was selected for its regular rectangular shape. We used three time-lapse cameras to monitor the flock, distributed on one side of the pasture. For the estimation of the animal location, $n_m = 30$ marks were distributed on the ground of the pasture. The pasture was virtually divided into three sub-pastures of approximately 20 m by 20 m, and each camera was responsible for monitoring one sub-pasture. The pasture was relatively flat, with no trees, but there was a concrete post in front of the second camera.

This data set will be referred to as D_1. A schematic representation of the pasture is presented in Fig. 2c. The marks and camera positions are also available in Supplementary Material S1.

2.2.1.2. Monitoring goats during pregnancy D_2.
In this experiment the set-up was similar, but no kids were present, as the goats were still pregnant and only one camera was used, for two consecutive days. The camera was set to take pictures from 6 a.m. to 6 p.m. but the image quality was insufficient during the afternoon because the sun was facing the camera. We only used the pictures from 6 a.m. to 2 p.m. for the analysis. No marks were distributed on the pasture as this data set was used to estimate animal detection efficacy only. The pasture is relatively flat, with one plastic tray in front of the camera and several patches of weeds on the opposite side of the pasture. As in the previous experiment, the monitored area is approximatively 20 m by 20 m. An example of a picture is available in Supplementary Material S2.

2.2.1.3. Analysis procedure.
To estimate the efficacy of the animal detection workflow, i.e., the combination of Yolo detection and image classification, we designed a graphical user interface that selects randomly one image and shows the bbox around the detected objects. Then the user was asked to click on the False Positive detections (FP), i.e. the detected objects that were not the studied animals, and on the False Negative detections (FN), i.e. the studied animals that were not detected. The True Positive (TP) detections were then automatically deduced. We proposed to quantify the detection workflow efficacy using the following criterions:
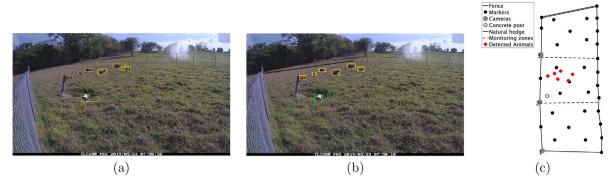
$$\text{Sensitivity} = \frac{TP}{TP + FN}.$$

$$\text{Precision} = \frac{TP}{TP + FP}.$$

The sensitivity is the percentage of detected animals and the precision is the percentage of correct detections. A sensitivity of 100% means that all the animals are always detected. A precision of 100% means that there are no false positive detections.

2.2.2. Spatial accuracy

In this section we present the determination of the accuracy of the animal location workflow, i.e. the combination of the projections of the centroids of the animals and the geometric transformation f, to estimate the spatial coordinates of the animals. The accuracy is the closeness of the estimated coordinates to the true locations.

We distinguished three different types of error: (i) the transformation error, (ii) the resolution error and (iii) the projection error. The transformation error is due to the use of the mapping function f. The accuracy of the mapping function varies with the number of marks that are used and with their spatial distribution. The form of the mapping function (e.g. polynomial, affine …) can also affect the accuracy. To estimate the value of the transformation error, we used one camera indoor to take a picture of a tile floor. In that the tile dimensions were known, it was easy to compute the spatial coordinates of each intersection of tiles, which can be used as a mark to build the set of Eq. (2). The floor was approximately 12 m by 7 m and we were able to

Fig. 2. Workflow of the goat monitoring. This is an example using data set D_1 and camera 2. (a) Original image with yellow bounding boxes (bboxes) around the objects detected by the neural network Yolo. (b) Detected objects after employing the non-max suppression algorithm to merge overlapping bboxes, and after image classification, to remove false positives (represented in green). The red points are the projected centroids of the detected objects (see Eq. (1)). The black line near the top of the image is the end of the monitoring area for the selected camera (here, camera 2). (c) Schematic representation of the monitored pasture with the estimated coordinates of the detected goats (red points). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

distinguish 285 marks. Some intersections were not used as they were not easily visible in the picture. An image of the experimental set-up is available in Supplementary Material S3.

We first studied the influence of the number of marks on the transformation error. We selected randomly a given number of marks to be used to estimate the mapping function (training data). Then, the true spatial coordinates of the unselected marks were compared to their estimated coordinates using the mapping function (test data). The operation was repeated 1,000 times in order to estimate the error distribution for a given number of marks, which we varied from 6 to 50. The true spatial coordinates of the marks were obtained by summing the number of tiles, adding an offset for the joints between the tiles. Finally, we used these results to select the optimal number of marks in order to estimate a model of the transformation error as a function of the distance to the camera.

The resolution error is related to the total number of pixels of the camera. Objects that are close to the camera appear bigger in the image and are thus described with a higher number of pixels than an object located far from the camera. If d_{xy} is the real world distance between two neighbouring pixels, then $d_{xy}^1 < d_{xy}^2$ if the first neighbouring pixels are closer to the camera than are the second neighbouring pixels. Then for the resolution error, the maximal accuracy is the maximal value of d_{xy} over all pairs of neighbouring pixels located in the study area. To estimate this accuracy, we used the same experimental set-up as for the transformation error. The mapping function was estimated using the 285 available marks and the function was then used to estimate the spatial coordinates of each pixel, which would be an approximation of their true spatial coordinates. We finally computed the distance between all the pairs of neighbouring pixels to find the maximal distance. We used these results to estimate a model of the resolution error as a function of the position of the object.

The projection error is related to the projection of the animal's centroid onto the ground of the pasture. Depending on the orientation and position of the animal, the projection of the centroid can be far from the true centroid, which produces an error in the animal's

location. To quantify this error, we used the data set D_1 to select randomly 200 detected animals for which we manually defined the projection of the animal's centroid onto the ground of the pasture. We then used the mapping function to compare the distance between the estimated projection point using Eq. (1) and the manually defined projection point.

3. Results

3.1. Efficacy of the automatic animal detection method

3.1.1. Data set D_1

About 6,200 pictures were analysed and an average sensitivity and precision of 79.5% and 90.5% were found. The results for each camera are presented in Table 1.

As is shown in Fig. 3, the detection errors are mainly concentrated in two particular zones (coloured in red in Fig. 3(a)). The first zone is behind a concrete post present in the pasture (see Fig. 2(a) and (b)). Around this post, the image category classifier barely discriminates the animals, which brings about a detection error. In the second zone, at the end of the pasture, there is a natural hedge that appears dark in the images. As a consequence, animals can be missed by Yolo, even more when the goats are black. 46% of the false positive errors and 53% of the false negative errors are inside these two zones. If they are removed from the analysis, the sensitivity and precision increase to 87.3% and 93.6%.

Detecting kids in pictures can be particularly challenging as they are generally very small. In some cases, they can be smaller than a weed, making them difficult to detect even though their heads are visible. As a result, 55% ([52%, 58.3%] 95% C.I. estimated on 1,000 images) of the false negative errors were due to missed kids.
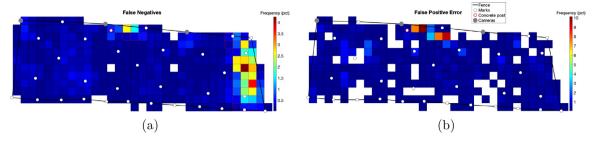
3.1.2. Data set D_2

In all, 500 random pictures were analysed and a sensitivity and precision of 94.8% and 92.5% were found. Compared to the previous

Table 1
Animal detection efficacy. *N* is the number of pictures used to estimate the precision (Prec.) and sensitivity (Sens.). Columns **All** are the results for the combined data sets.

	Dataset D_1			Dataset D_2			All		
Cameras	*N*	Prec.	Sens.	*N*	Prec.	Sens.	*N*	Prec.	Sens.
1	1,813	90.8%	84.3%	500	92.5%	94.8%	6,550	90.75%	84.6%
2	2,252	83.8%	88.6%						
3	1,985	95.6%	70.7%						

Fig. 3. Spatial distribution of the detection error for data set D_1. Spatial distribution of the false negative detections (a) and false positive detections (b). The red areas in (a) are the areas where the detection errors are concentrated. 46% of the false positive errors and 53% of the false negative errors are inside the red areas. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

data set, there is a smaller FN error rate, which increases the sensitivity. This is probably due to the fact that no objects were hiding the animals and that no kids were present in the pasture. Many, if not all, of the FP errors were due to the different patches of weeds and to the plastic tray.

3.2. Efficacy of the animal location estimation method

3.2.1. Transformation error
3.2.1.1. Number of marks.
The distribution of the transformation error as a function of the number of marks is presented in Fig. 4(a). There is a high error variance, so that the average error and median are significantly different when fewer than 16 marks are used. This means that the transformation error can be greatly affected by the spatial distribution of the marks (selected randomly in our case). One can see that the transformation error decreases as the number of marks increases, although the error becomes approximately stable when more than 22 marks are used. With 22 marks, the average error is equal to 12.8 cm.

3.2.1.2. Transformation error model.
Based on the previous results, we used the transformation error when 25 marks were used. We then fitted a polynomial model of the transformation error of the form

$$T_e = p_1 \times d^2 + p_2 \times d + p_3,$$

where d is the real-world distance between the object and the camera and $(p_1, p_2, p_3) = (2.873e^{-5}, -0.0301, 13.67)$. We used a robust bi-square linear least-squares fitting method to estimate the parameters of the model ($R^2 = 0.72$).

3.2.2. Resolution error
The resolution error is equal to 0.7 cm, 1.1 cm, 2 cm and 4.1 cm when the object is at a real-world distance of 5 m, 10 m, 15 m and 20 m. We fitted the following model for the resolution error:

$$T_r = \alpha + \beta_1 x + \beta_2 y + \beta_{12} xy + \beta_{11} x^2 + \beta_{22} y^2,$$

where (x, y) are the spatial coordinates of the object and

$$(\alpha, \beta_1, \beta_2, \beta_{12}, \beta_{11}, \beta_{22})$$
$$= (0.63, -5.93e^{-4}, 1.42e^{-3}, -5.17e^{-7}, -1.64e^{-7}, 7.57e^{-7}).$$

We used a robust bi-square linear least-squares fitting method to estimate the parameters of the model ($R^2 = 1$).
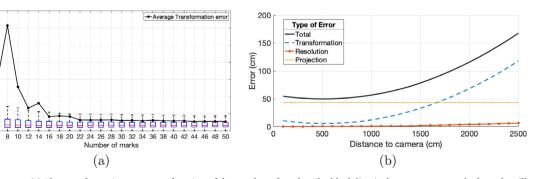
3.2.3. Projection error
The average projection error was estimated from 200 random pictures selected randomly. It is equal to 43.39 cm.

3.2.4. Total location error
The total location error is provided in Fig. 4(b). The maximal accuracy of the method is equal to 56 cm, 78.5 cm, 115.9 cm and 167.8 cm when the object is located at a real-world distance of 10 m, 15 m, 20 m and 25 m.

4. Discussion

We have presented a framework for animal tracking at the pasture scale combining time-lapse cameras, deep learning, and image registration. This study showed that using computer vision for animal tracking is also possible in outdoor conditions, and not just indoors with top view cameras. We estimated the sensitivity of the method to be nearly 85%, which is lower than indoor studies. For example, in (Kashiha et al., 2014), 10 pigs were tracked in a 2.25 m × 3.60 m pen with a sensitivity of nearly 90%. In (Nasirahmadi et al., 2016), 22–23 pigs were tracked in a 6.75 m × 3.10 m pen with a sensitivity of nearly 95%. But one has to take into consideration that in these cases, the lighting conditions and camera angle are more favorable, while the study area and the number of animals were smaller. The efficacy of our framework is improved when no objects other than animals are present inside the pasture and there is no background with a colour that can be confused with the colour of an animal. Patches of ungrazed weeds could

Fig. 4. Animal location error. (a) The transformation error as a function of the number of marks. The black line is the average error, the box-plots illustrate the error distribution. The red line is the median, the top and bottom blue lines are the 75th and 25th percentiles, the black lines are the whiskers, not considered as outliers. For clarity, we do not show the outliers, as some of them were far from the median. (b) Model of the different types of error as functions of the object's real-world distance from the camera. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

also be removed before the monitoring in order to improve the efficacy, although it can change animal behaviour. The relative size of the animals compared to the size of the weeds also plays an important role in the efficacy of the framework. The relative location of the camera compared to the sun can also influence the efficacy of the framework. In our case, we were not able to analyse the afternoon images for data set D_2 because the sun was facing the camera. For locating the animals, the maximal accuracy of the method is around 115 cm for object located at distances of up to 20 m, which is probably the maximal monitoring distance with the tested cameras. This accuracy is as good or better than most GPS (Buerkert and Schlecht, 2009; Benvenutti et al., 2015). When the object is located less than 10 m from the camera, most of the location error is due to the projection error, i.e. the accuracy of the projection of the goat's centroid. In our case, we used a simple equation to define the projection, but a more sophisticated model can be used, as for example a model that could take into account the position/orientation of the animal and/or the dimensions of the bbox. When the object is located more than 10 m from the camera, the transformation error is responsible for more than 22% of the total location error, but the projection error is still the highest source of error until 17 m, where the transformation error becomes the highest source of error. At any distance, the resolution error is nearly negligible.

The animal detection method could be improved in several ways. First, cameras with higher quality images can be used, although the resulting gains have to be balanced with the battery life. The object classification step could also be improved, for example, by using a larger training set. The object classification could also be done with more than the two classes *Goat* and *No Goat*. For data set D_2 for example, efficacy might be improved by using a class for the patches of ungrazed weeds, one class for the plastic tray, and one class for the goats. Furthermore, a dedicated CNN could be used to detect farm animals in the images, which could remove the need for the object classification step. In the future, it could be valuable to collectively assemble a large training data-set dedicated to farm animals.

5. Conclusions

Image analysis combined with time-lapse cameras can be a valuable alternative to GPS collars. The efficacy of the technique depends on the size of the animal, the locations of the cameras, and the set-up of the pasture.

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

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.compag.2019.105150.

References

Adrion, F., Kapun, A., Eckert, F., Holland, E.M., Staiger, M., Götz, S., Gallmann, E., 2018. Monitoring trough visits of growing-finishing pigs with uhf-rfid. Comput. Electron. Agric. 144, 144–153. https://www.sciencedirect.com/science/article/pii/S0168169917300157.

Benvenutti, M., Coates, T., Imaz, A., Flesch, T., Hill, J., Charmley, E., Hepworth, G., Chen, D., 2015. The use of image analysis to determine the number and position of cattle at a water point. Comput. Electron. Agric. 118, 24–27. https://www.sciencedirect.com/science/article/pii/S0168169915002422.

Brünger, J., Traulsen, I., Koch, R., 2018. Model-based detection of pigs in images under sub-optimal conditions. Comput. Electron. Agric. 152, 59–63. https://www.sciencedirect.com/science/article/pii/S0168169917313741.

Buerkert, A., Schlecht, E., 2009. Performance of three gps collars to monitor goats' grazing itineraries on mountain pastures. Comput. Electron. Agric. 65, 85–92. https://doi.org/10.1016/j.compag.2008.07.010. http://www.sciencedirect.com/science/article/pii/S0168169908001816.

Fretwell, P.T., Staniland, I.J., Forcada, J., 2014. Whales from space: counting southern right whales by satellite. PLoS One 9, e88655.

Fukunaga, T., Kubota, S., Oda, S., Iwasaki, W., 2015. Grouptracker: video tracking system for multiple animals under severe occlusion. Comput. Biol. Chem. 57, 39–45. https://www.sciencedirect.com/science/article/pii/S1476927115000237.

González, L., Tolkamp, B., Coffey, M., Ferret, A., Kyriazakis, I., 2008. Changes in feeding behavior as possible indicators for the automatic monitoring of health disorders in dairy cows. J. Dairy Sci. 91, 1017–1028.

Goshtasby, A., 1988. Image registration by local approximation methods. Image Vis. Comput. 6, 255–261. https://www.sciencedirect.com/science/article/pii/0262885688900169.

Hollings, T., Burgman, M., Van Andel, M., Gilbert, M., Robinson, T., Robinson, A., 2018. How do you find the green sheep? a critical review of the use of remotely sensed imagery to detect and count animals. Methods Ecol. Evol. 9, 881–892. https://doi.org/10.1111/2041-210X.12973.

Kamilaris, A., Prenafeta-Boldú, F.X., 2018. Deep learning in agriculture: A survey. Comput. Electron. Agric. 147, 70–90. https://www.sciencedirect.com/science/article/pii/S0168169917308803.

Kashiha, M.A., Bahr, C., Ott, S., Moons, C.P., Niewold, T.A., Tuyttens, F., Berckmans, D., 2014. Automatic monitoring of pig locomotion using image analysis. Livestock Sci. 159, 141–148. https://www.sciencedirect.com/science/article/pii/S1871141313005003.

Kolarevic, J., Aas-Hansen, Ø., Espmark, Å., Baeverfjord, G., Terjesen, B.F., Damsgård, B., 2016. The use of acoustic acceleration transmitter tags for monitoring of atlantic salmon swimming activity in recirculating aquaculture systems (ras). Aquacul. Eng. 72, 30–39. https://www.sciencedirect.com/science/article/pii/S0144860916300450.

Kongsro, J., 2014. Estimation of pig weight using a microsoft kinect prototype imaging system. Comput. Electron. Agric. 109, 32–35. https://doi.org/10.1016/j.compag.2014.08.008. http://www.sciencedirect.com/science/article/pii/S0168169914002075.

Leonard, S.M., Xin, H., Brown-Brandl, T., Ramirez, B.C., 2019. Development and application of an image acquisition system for characterizing sow behaviors in farrowing stalls. Comput. Electron. Agric. 163, 104866. https://www.sciencedirect.com/science/article/pii/S0168169919305666.

Mahajan, S., Dogra, R., 2015. A review on image enhancement techniques. Int. J. Eng. Innovat. Technol. 4, 108–113. https://pdfs.semanticscholar.org/957a/f522302f18b0d2d818a675c58954987c656b.pdf.

Mortensen, A.K., Lisouski, P., Ahrendt, P., 2016. Weight prediction of broiler chickens using 3d computer vision. Comput. Electron. Agric. 123, 319–326.

Nasirahmadi, A., Edwards, S.A., Sturm, B., 2017. Implementation of machine vision for detecting behaviour of cattle and pigs. Livestock Sci. 202, 25–38. https://doi.org/10.1016/j.livsci.2017.05.014. http://www.sciencedirect.com/science/article/pii/S1871141317301543.

Nasirahmadi, A., Hensel, O., Edwards, S.A., Sturm, B., 2016. Automatic detection of mounting behaviours among pigs using image analysis. Comput. Electron. Agric. 124, 295–302.

Pal, N.R., Pal, S.K., 1993. A review on image segmentation techniques. Pattern Recognit. 26, 1277–1294. https://pdfs.semanticscholar.org/12e8/ac67332bddc144527179ce48d9df07deb609.pdf.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788.

Swain, D.L., Friend, M.A., Bishop-Hurley, G., Handcock, R., Wark, T., 2011. Tracking livestock using global positioning systems–are we still lost? Animal Product. Sci. 51, 167–175.

Turner, L., Udal, M., Larson, B., Shearer, S., 2000. Monitoring cattle behavior and pasture use with gps and gis. Can. J. Animal Sci. 80, 405–413.

Vayssade, J.A., Arquet, R., Bonneau, M., 2019. Automatic activity tracking of goats using drone camera. Comput. Electron. Agric. 162, 767–772. https://www.sciencedirect.com/science/article/pii/S0168169918312894.

Villa, A.G., Salazar, A., Vargas, F., 2017. Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks. Ecol. Informat. 41, 24–32. https://www.sciencedirect.com/science/article/pii/S1574954116302047.

Zhou, C., Zhang, B., Lin, K., Xu, D., Chen, C., Yang, X., Sun, C., 2017. Near-infrared imaging to quantify the feeding behavior of fish in aquaculture. Comput. Electron. Agric. 135, 233–241. https://www.sciencedirect.com/science/article/pii/S0168169916311930.

Zitova, B., Flusser, J., 2003. Image registration methods: a survey. Image Vision Comput. 21, 977–1000. https://www.sciencedirect.com/science/article/pii/S0262885603001379.