Survey of Livestock Counting and Tracking Methods

Farah Sarwar 1 *, Timotius Pasang 1, Anthony Griffin 1, Saeed-Ur-Rehman 2

1Auckland University of Technology, 55 Wellesley Street East, Auckland, 1142, New Zealand
2Flinders University, Sturt Rd, Bedford Park, 5042, South Australia

*Corresponding author:
E-mail: farah.sarwar@aut.ac.nz

ABSTRACT

Conventional livestock counting methods are tedious, time-consuming, and labor-intensive for farmers, which makes counting an irregular task. This inability to constantly monitor stock numbers gives farm rustlers sufficient time to steal farm animals and hence leads to a significant financial loss annually. To overcome this issue, research using unmanned aerial vehicles (UAVs) in pastoral farming is growing progressively since the last decade, but their use is still limited to some extent. This research article gives a detailed analysis of the existing hardware-based and software-based methods. The impact of shifting from current methods to a UAV-based system is discussed using the findings of research articles, and algorithms for object detection in images and tracking in videos are also analyzed. The article concludes that there are still unexplored practical uses of UAV in pastoral farming for monitoring, counting, and tracking farm animals, especially in countries like New Zealand, where pastoral products cover a major portion of export revenue.

Keywords: Livestock, counting, UAV, deep learning

Introduction

Fast pasture growth makes dairy farming low cost, sustainable, and efficient for New Zealand to compete well as an exporter of food and fiber. According to the statistical survey of New Zealand for the year 2016, 55,473 farms have an average area of 252 hectares each and 44% are of sheep or beef, 21% are dairy, and 6% are mixed farms (Beef and Lamb New Zealand, 2019). Sheep revenue alone contributes 48% to the gross farm revenue and total pastoral products covered almost 44% of the actual export revenue of New Zealand in 2017. For the year ended 30 September 2017, approximately 19.5 million lambs, 3.7 million sheep, and 2.4 million cattle are processed to produce 362,000, 94,000, and 633,000 tons of meat, respectively. While deer and cattle farming has increased in the last years, sheep farming has declined by 0.8% (Beef and Lamb New Zealand, 2019). The reason behind this decline is the requirement of extra labor and effort in handling sheep farms.

With the advancement in pastoral and agricultural techniques, a variety of data is available to farmers, such as expected climatic changes, area temperatures, soil dampness values, pasture covers, and individual stock identification for growth rates monitoring. Yet most of the tasks done in livestock handling are labor-intensive, and counting paddock animals is one of them. Counting is usually done only fewer times a year and thus makes stock information infrequent. It also seems disruptive for the farm animals because they have to pass through narrow choke points for this purpose. There are many cases where sheep and especially pregnant ewes were stolen from farms, and farmers found out about their financial loss after two to three months in the next sheep count (Johnstone, 2012). Easy and robust remote livestock counting and monitoring is one such system for handling the herds efficiently and can also help to respond to any disturbance in livestock.

How to cite:
Sarwar, F., Pasang, T., Griffin, A., Rehman, S.U. (2020). Survey of livestock counting and tracking methods. 1st International Conference Eco-Innovation in Science, Engineering, and Technology. NST Proceedings. pages 150-157. doi: 10.11594/ nstp.2020.0523
promptly. The main purpose of this article is to highlight the existing methods for livestock monitoring, counting, and tracking, and to identify the main barriers in introducing new methods in this field. If more sophisticated ways for these tasks can be designed, it will benefit both farmers and farm animals.

**Literature Review**

This section discusses the existing literature and a comprehensive study has been carried out to provide critical analysis of all methods. Current literature can be broadly classified into two main categories: hardware-based methodologies and software-based methodologies. These categories are discussed briefly in the following subsections.

**Hardware-based methodologies**

The existing methods for animal identification and counting comprise radiofrequency identification (RFID) bolus, RFID ear tags (Kassler, 2001), electronic identification (EID) ear tags, and GPS collars (Handcock et al., 2009). RF bolus is injected into the animal bolus to track it throughout its life. Though it seems a one-time process, these highly technological items make it costly to use them for every purchased animal on the farm. RF ear tags are available in a large variety and a few countries can be reused after reprinting. The EID tags work similarly to RFID ear tags with the difference of having an electronic chip in them. These methods need a direct encounter with farm animals and are comparatively disruptive for them (Kassler, 2001). However, these methods do not give automatic stock count information to the farmers and are not convenient in terms of time and cost.

An expensive approach to get quick and automatic herd count is the use of GPS devices (Handcock et al., 2009) that are usually used as a collar, and can track the location. Though they also offer the benefit of health monitoring, the minimum cost per collar is US$60. Due to this reason, these collars are difficult to use even in farms of a few hundred animals and are suitable mostly for cows and buffalos. Maintenance of these collars in terms of replacing batteries can increase the involvement of farmers instead of reducing it.

The use of UAV for the same task is another option that can be effectively utilized. There is a large variety of drones, ranging from the ones that can be used for recreational purposes, with lower prices, to those that are good for professional use and are highly-priced. Researchers usually prefer to use UAV with functions like GPS-enable autopilots, emergency landing technology, stability, good camera resolution, long battery life, and easy flight mode. It gives ease in access to the images and videos and can be easily linked with smartphones or tablets. Also, the latest UAV can go up to a range of 7 km from the control box and have a flight time of 25 minutes. Civil aviation authority of countries like New Zealand allows UAV flight in the line of sight. Currently, UAVs are in use in various domains such as remote security, wildlife monitoring (Gonzalez et al., 2016), livestock detection (Sarwar et al., 2018), livestock counting (Sarwar et al., 2020), surveillance, security and many more. This system offers a one-time investment and provides the required stock information soon after recording the paddock footage. Figure 1 shows RFID tags, RFID bolus, GPS collar, and a UAV, that are easily available in the market.
Software-based methodologies

To process the videos recorded by UAV for animal monitoring, counting, and tracking, machine learning, or deep learning-based strategies are used. Machine learning is defined as a science of techniques and approaches for problem-solving of artificially intelligent systems, whereas, deep learning is an inspiration of the human brain, and use multi-layered neurons for data processing in an effective way. Though deep learning needs an enormous amount of data and very efficient high-end machines with discrete graphics cards for processing, once trained, it can provide results very fast. The main difference between these approaches is that the latter performs feature extraction and classification as an end-to-end process. Existing machine learning and deep learning algorithms for object detection and tracking in videos are discussed below very briefly. The researcher can select the algorithm depending on the approach used for animal counting in the paddock. Object detection can be used if only images are collected and object tracking can be used if videos are recorded.

Object detection in images

Handcrafted features extraction for object classification and detection were introduced in the late 2000s. It laid the foundation for techniques which are scale, translation, rotation, illumination, and viewpoint invariant. It started with the introduction of scale-invariant feature transform (SIFT) (Lowe, 1999), and support vector machine (SVM) (Suykens & Vandewalle, 1999). Researchers found a new way of handling object detection problems using features, and many algorithms were proposed in the same domain. Some of the prominent contributions are cascade networks (Viola & Jones, 2001), a bag of words (Sivic & Zisserman, 2003), a histogram of gradients (HOG) (Dalal & Triggs, 2005), spatial pyramid matching (SPM) (Lazebnik, Schmid, & Ponce, 2006), deformable part-based model (DPM) (Felzenszwalb, McAllester, & Ramanan, 2008), HOG based local binary patterns (HOG-LBP) (Wang et al., 2009), improved fisher kernels (Perronnin, Sanchez, & Mensink, 2010) and selective search (Van de Sande et al., 2011).

The research took an important turn with the introduction of convolutional neural networks (CNNs) for object detection in images (Krizhevsky, Sutskever, & Hinton, 2012) accomplished the classification task on ImageNet using deep CNN (DCNN), AlexNet. After seeing promising results by CNNs, researchers starting exploring the impact of deeper networks and different ways of training them. According to Girshick et al. (2014) proposed a Region-based CNN (R-CNN) algorithm and Fast R-CNN (Girshick, 2015) for the same network but comparatively better results. Ren et al. (Ren, Girshick, & Sun, 2015) designed an even faster algorithm, Faster R-CNN, with a bit higher efficiency, and then new deeper networks and faster algorithms were proposed by many
other researchers; you only look once (YOLO) (Redmon et al., 2016, Redmon and Farhadi, 2017), single-shot multibox detector (SSD) (Liu et al., 2016), Mask R-CNN (He et al., 2017), RetinaNet with focal loss (Lin et al., 2017) and U-Net with weighted Hausdorff distance (WHD) (Ribera et al., 2018). Multiple versions of YOLO are available and are the fastest available algorithms for video processing that can be used for real-time processing. Figure 2 shows a brief timeline of object detection methods.

Figure 2. Few milestones for object detection using (a) machine learning and (b) deep learning algorithms

The prospective, that livestock monitoring can help farmers in identifying animal-related issues, has motivated various researchers during the last decade to analyze the videos recorded in different areas and provide required information to the concerned persons after processing them. The research to design wildlife monitoring methods (Jachmann, 1991), who put effort into estimating elephant density estimation through aerial images. Some of the previously used techniques in wildlife life monitoring are AdaBoost classifier (Burghardt and Calic, 2006), power spectral-based techniques (Parikh, Patel, & Bhatt, 2013), DPM, SVM (van Gemert et al., 2014), DCNN (Sarwar et al., 2020) and the template matching algorithm (Gonzalez et al., 2016).

**Object tracking in videos**

Object tracking in a video can be defined as the approximation of the path of an object throughout the video by keeping track of objects’ spatial changes like variations in position, size, and shape. Like detection, object tracking algorithms are also classified into two domains, machine learning, and deep learning. Machine learning algorithms are further classified into three main categories: point tracking, kernel-based tracking, and silhouette based tracking.

Point-based tracking involves object detection per frame using different feature points. If the object is detected successfully in an image, then these algorithms have a high rate of accuracy, however, can lead to false object detections in case of occlusions. A few main algorithms that fall in this category are Kalman filter (Kalman, 1960) and particle filter (Kitagawa, 1987).

Kernel-based tracking tracks the motion of the object from one frame to the next frame using the primary object’s region. The algorithms that come under this domain are categorized depending upon the method that is used to track the object. A few of the most commonly used algorithms
are template matching, layering, mean shift (Exner et al., 2010), and SVM (Suykens & Vandewalle, 1999).

Silhouette-based tracking algorithms are useful for tracking objects with complex shapes like hands and fingers, that cannot be described completely using simple geometric features. These algorithms use rough shape models of objects from the previous frame to track them in the preceding frame. They use predefined information for each object and handle different complex shaped objects, occlusion, merging, and splitting problems. The main subcategories are contour and shape tracking (Parekh, Thakore, & Jaliya, 2014). The most important advantage of tracking silhouettes is its flexibility to handle a large variety of object shapes. However, they need prior information of all the objects under consideration.

Deep learning algorithms are used solely and also in conjunction with machine learning in this area. Lijun Wang et al. (Wang et al., 2015) used the features of a fully convolutional network to track the objects in a video. Although they tracked a single object per video, they were able to overcome many tracking challenges. According to Kang et al. (2016) combined tubelet, bounding box sequences, and CNNs (T-CNN) for the same task, where low-confidence detection classes were suppressed to reduce false positives, and the detection results were propagated between adjacent frames for reducing false negatives. According to Feichtenhofer, Pinz and Zisserman (2017) proposed a ConvNet architecture to detect and track the objects in a video jointly, hence, named it the Detect and Track (D&T) approach. They trained a fully end-to-end convolutional network using video frames and after computing cross-correlation between feature responses of adjacent frames, region of interest (ROI) pooling layer for classification and regression of proposal boxes is used. Table 1 shows a comparison of the discussed techniques for object tracking.

**Table 1. Comparison of the tracking algorithm**

| Tracking Algorithms | Types | Reference no. | Advantages | Limitations |
|--------------------|-------|---------------|------------|-------------|
| **Point Tracking** | Kalman Filter | (Kalman, 1960) | Can track objects in noisy images | Assumes Gaussian distribution of all state variables |
| | Particle Filter | (Kitagawa, 1987) | Does not need Gaussian distribution of state variables | Difficulty in tracking multiple objects and computationally expensive |
| | Template Matching | (Parikh et al., 2013) | Relatively simple method and can deal with partial occlusion | High computational cost |
| | Mean shift | (Exner et al., 2010) | Does not need any prior shape information | Cannot distinguish an object from the background in case of the same color |

To be continue
**Table**

| Method                        | Description                                                                 | Requirements                                      |
|-------------------------------|-----------------------------------------------------------------------------|---------------------------------------------------|
| SVM                           | Suitable for multiple object tracking                                        | Needs training                                    |
| Layering Based                | Suitable for multiple object tracking                                        | High computational cost                           |
| Contour Matching              | Can handle multiple complex-shaped objects                                    | Computationally expensive                         |
| Shape Matching                | Resistant to noise and can handle multiple complex-shaped objects             | Need prior information of all shapes and needs training |
| T-CNN                         | Low computational cost and good for tracking a few objects                    | Difficulty in tracking many objects in video      |
| Deep CNN Based Tracking       | Failed detections can be recovered using previous frames detections          | Computationally expensive algorithm                |
| FCN                           | Can handle occlusion and noise                                                | Track one object at a time                        |

**Discussion**

Existing hardware-based methods ultimately lead to direct farmer-animal encounters and can be categorized as manual counting methods. Though machines are available that can be installed near holding pens for automatic tag reading, the gathering of full stock for daily information is not feasible. Manual counting of the livestock can be time-consuming and prone to various psychological phenomena that can lead to bias or optical illusion. Good use of technology is required for the betterment of farmers. Surveillance or motion-sensor cameras can be installed around the whole farm, however, having an average paddock of 252 hectares, many cameras will be needed for proper coverage. There will be the problems of continuous power supply, networking issues to link all these cameras together and data transmission to a central system. It may cost a lot more than the annual financial loss.

These factors support automatic detection, counting, and remote monitoring of livestock using a UAV. This would give much relief to farmers and extract relevant information very quickly. However, the design of this automatic system is not trivial, even if the data is collected under the most favorable conditions. Illumination, background, and shadow variations offer many challenges. Thus finding a general technique to handle these issues, while incorporating different farms, presents many technical issues. The main challenge, which UAV offers for this task, is the coverage of the whole paddock from a good allowable height and in minimum possible time. There is a great variation in paddocks as few paddocks have plain ground with no trees while others have uneven terrains and many bushes. Currently, no large data set is publically available, so, creating a good and large dataset will be one of the main contributions towards this research. Figure 3 shows few sheep farm images taken by a UAV, where detected sheep has a blue dot on it and
looks like a whitish blob from 80 m altitude. Figure 3, a sub-image of full frame, gives a hint of another challenge, to detect and track hundreds of animals per each video frame, where each animal covers only 10s of pixels and looks like a colored blob from a good height.

![Figure 3. Sheep detected by the U-Net (Sarwar et al., 2020) in a video recorded by a UAV at 80 m height](image)

**Conclusion**

The RFID tags and bolus for livestock detection, monitoring, and tracking are in use for decades, however, the use of a UAV is still in its infancy. They discussed literature shows the potential, benefits, and challenges of using a UAV, along with object detection and tracking algorithms, for these tasks. There is still a lot of room available for research in this area and researchers can explore the algorithms that can detect and track livestock from different altitudes, and the higher the better and more time-efficient the system will be. It is a new direction of research and with the improvement in the UAV technology and increased efficiency of algorithms, farmers will be able to reduce the burden of many manual tasks in pastoral farming.

**References**

Beef and Lamb New Zealand. (2019). *New Season Outlook 2018-19* [Online]. Available: [https://beeflambnz.com/sites/default/files/news-docs/WGTN_DOCS-%23185883-v3-P18017_New_Seaons_Outlook_2018-19_-_Report%255b1%255d.pdf](https://beeflambnz.com/sites/default/files/news-docs/WGTN_DOCS-%23185883-v3-P18017_New_Seaons_Outlook_2018-19_-_Report%255b1%255d.pdf) [Accessed 26 July 2019].

Burghardt, T. & Calic, J. (2006). Real-time face detection and tracking of animals. *8th IEEE Seminar on Neural Network Applications in Electrical Engineering*, 291, 27-32.

Dalal, N. & Triggs, B. (2005). Histograms of oriented gradients for human detection. *Proceedings of the IEEE Computer Society Conference on CVPR*, 1, 886-893.

Exner, D., Bruns, E., Kurz, D., Grundhöfer, A. & Bimber, O. (2010). Fast and robust CAMShift tracking. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 9-16.

Feichtenhofer, C., Pinz, A. & Zisserman, A. (2017). Detect to track and track to detect. *Proceedings of the IEEE Conference on CVPR*, 3038-3046.

Felzenszwalb, P. F., McAllester, D. A. & Ramanan, D. (2008). A discriminatively trained, multiscale, deformable part model. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2, 1-8.

Girshick, R. (2015). Fast R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*, 1440-1448.

Girshick, R., Donahue, J., Darrell, T. & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE conference on CVPR*, 580-587.

Gonzalez, L., Montes, G., Puig, E., Johnson, S., Mengersen, K. & Gaston, K. (2016). *Unmanned aerial vehicles (UAVs) and artificial intelligence revolutionizing wildlife monitoring and conservation*. Sensors, 16, 97.

Handcock, R., Swain, D., Bishop-Hurley, G., Patison, K., Wark, T., Valencia, P., Corke, P. & O’Neill, C. (2009). Monitoring animal behaviour and environmental interactions using wireless sensor networks, GPS collars and satellite remote sensing. *Sensors*, 9, 3586-3603.

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

Jachmann, H. (1991). Evaluation of four survey methods for estimating elephant densities. *African Journal of Ecology*, 29, 188-195.
Johnstone, T. (2012, March 31). Farmer appeals for help over sheep theft [Online]. Available: https://www.nzherald.co.nz/wairarapa-times-age/news/article.cfm?c-id=1503414&objectid=11058889 [Accessed 2 June 2019].
Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering, 82*, 35-45.
Kang, K. Li, H., Yan, J., Zeng, X., Yang, B., Xiao, T., Zhang, C., Wang, Z., Wang, R. & Wang, X. (2016). T-CNN: Tubelets with convolutional neural networks for object detection from videos. *arXiv preprint arXiv:1604.02532*.
Kassler, M. (2001). LIVE. 106 automatic counting of sheep [Online]. Australia: Meat & Livestock Australia Ltd. Available: http://www.livcorp.com.au/LC/files/59/59d3f032-ebca-4ff2-bcd3-16f8f9b83e0b.pdf [Accessed 5 February 2019].
Kitagawa, G. (1987). Non-gaussian state—Space modeling of nonstationary time series. *Journal of the American Statistical Association, 82*, 1032-1041.
Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2012). IMagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems, 1097-1105*.
Lazebnik, S., Schmid, C. & Ponce, J. (2006). Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. *Proceedings of the IEEE Computer Society Conference on CVPR*, 2, 2169-2178.
Lin, T.-Y., Goyal, P., Girshick, R., He, K. & Dollár, P. (2017). Focal loss for dense object detection. *Proceedings of the IEEE International Conference on Computer Vision, 2980-2988*.
Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y. & Berg, A. C. (2016). SSD: Single shot multibox detector. *European Conference on Computer Vision, 21-37*.
Lowe, D. G. (1999). Object recognition from local scale-invariant features. *International Conference on Computer Vision, 99*, 1150-1157.
Parekh, H. S., Thakore, D. G. & Jaliya, U. K. (2014). A survey on object detection and tracking methods. *International Journal of Innovative Research in Computer and Communication Engineering, 2*, 2970-2978.
Parikh, M., Patel, M. & Bhatt, D. (2013). Animal detection using template matching algorithm. *International Journal of Research in Modern Engineering and Emerging Technology, 1*, 26-32.
Patel, C. I. & Patel, R. (2012). Contour based object tracking. *International Journal of Computer and Electrical Engineering, 4*, 525-528.
Perronnin, F., Sánchez, J. & Mensink, T. (2010). Improving the fisher kernel for large-scale image classification. *European Conference on Computer Vision, 143-156*.
Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 779-788*.
Redmon, J. & Farhadi, A. (2017). YOLO9000: better, faster, stronger. *Proceedings of the IEEE Conference on CVPR, 7263-7271*.
Ren, S., He, K., Girshick, R. & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems, 91-99*.
Ribera, J., Güera, D., Chen, Y. & Delp, E. (2018). Weighted Hausdorff distance: A loss function for object localization. *arXiv preprint arXiv:1806.07564*.
Sarwar, F., Griffin, A., Periasamy, P., Portas, K. & Law, J. (Year). Published. Detecting and counting sheep with a convolutional neural network. *Proceedings of the 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2018. 1-6.
Sarwar, F., Griffin, A., Rehman, S. U. & Pasang, T. (2020). Towards Detection of Sheep Onboard a UAV. *arXiv preprint arXiv:2004.02758*.
Sivic, J. & Zisserman, A. (2003). Video Google: A text retrieval approach to object matching in videos. *Proceedings of the IEEE International Conference on Computer Vision, 1*, 1-8.
Suykens, J. A. & Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural processing letters, 9*, 293-300.
Van de Sande, K. E., Uijlings, J. R., Gevers, T. & Smeulders, A. W. (2011). Segmentation as selective search for object recognition. *IEEE International Conference on Computer Vision (ICCV)*, 1, 1879-1886.
van Gemert, J. C., Verschoor, C. R., Mettes, P., Epema, K., Koh, L. P. & Wich, S. (2014). Nature conservation drones for automatic localisation and counting of animals. *European Conference on Computer Vision, 255-270*.
Viola, P. & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1*, 511-518.
Wang, L., Ouyang, W., Wang, X. & Lu, H. (2015). Visual tracking with fully convolutional networks. *Proceedings of the IEEE International Conference on Computer Vision, 31*, 3119-3127.
Wang, X., Han, T. X. & Yan, S. (2009). An HOG-LBP human detector with partial occlusion handling. *Proceedings of the 12th international IEEE Conference on Computer Vision, 32-39*.