REAL-TIME OPEN FIELD CATTLE MONITORING BY DRONE: A 3D VISUALIZATION APPROACH

Fei Yang, Ningbo Zhu, Shuonan Pei and Irene Cheng

University of Alberta, Computing Science Department, Multimedia Program
2-32 Athabasca Hall, University of Alberta, Edmonton, Alberta, T6G 2E8, Canada

ABSTRACT
Monitoring a large herd across an open field is challenging in the agricultural industry but is essential for the welfare of cattle. With the advancement of Unmanned Aerial Vehicle (UAV) technology, drones are now commonly used for surveillance. In this work, we apply UAV technology and drone-captured videos to monitor cattle in open pastures. We use cow headcount as a use case. Although cattle headcount in a confined indoor environment has been studied extensively, our contribution lies in developing a framework that can localize the cows in the video and track their movements on a 3D canvas in real-time. By using a 3D visualization approach, we expect to resolve many of the occlusion issues by guiding the drone operator to navigate the drone to discover important information. We use a pre-trained Mask R-CNN classification model to detect and track cows in the video. We then use Matplotlib 3D to create the 3D canvas and display the relative cow positions. Our real-time 3D cow visualization framework allows tracking herds remotely, saving time and labor for on-site manual herding, as well as providing better global monitoring of the herd. The complete implementation can be found in our publicly available GitHub link upon request.

KEYWORDS
Object Detection, Image Processing, Drone Video Analysis, 3D Visualization, Remote Monitoring, Livestock Management

1. INTRODUCTION
Governments maintain cattle welfare standards, which the agriculture sector has to follow. Keeping the headcount of livestock is a part of the process but it is a challenging task when herds are scattered in open fields. It is labor intensive and time consuming when on-site manual cattle-herding is performed. It can be expensive, as it may require farmers to slog through pasture, bush, wind, and snow in order to monitor and search them. To keep track of headcount too close to the cattle, i.e., cows in this work, often creates accidents especially when the cows are under stress. Every year, many people are hurt by cattle, mostly when cattle kick or crush them, since cattle have a weight advantage over human and cattle can move surprisingly fast. Agitated cattle are of a particular risk, and it takes training to handle them safely. Furthermore, manual headcount usually takes multiple trials to get it right because cows move, hide behind each other, or stand behind obstacles, e.g., a big hay bale or a tree. Counting is difficult even when the cattle are lying down.

According to the livestock statistics from the Government of Alberta, there were over 4.9 million units of cattle and 1.6 million pigs in 2020. These are staggering numbers. Hence, drone technology is becoming popular to provide smart agriculture, as it can find different applications here, including perform headcount of livestock across large open grazing areas. As the technology evolves, the costs of drones have come down and the accessibility and utility of drones have improved. Yinka and Ajayi (2019) introduced the different types of Unmanned Aerial Vehicles (UAVs) and the applications of UAVs in crop farming and in livestock. They also discussed some of the open challenges to the application of UAVs in Agriculture. Maddikunta et al. (2021) explored and presented the adaptation and usage of UAVs, and identified the key requirements of UAVs in smart agriculture, such as the accuracy of results, network availability, and data storage. Xu et al. (2020) presented a system that combined a quadcopter and artificial intelligence image processing for automated livestock detection. They applied the Mask R-CNN algorithm and demonstrated its effectiveness in livestock monitoring task. Barbedo and Koenigkan (2018) found that there were a limited number of UAV
applications used to monitor and count cattle. They analyzed the reasons for this apparent lack of progress and discussed both the technical challenges and the difficulties in finding target users. Hsieh et al. (2017) created a large drone view dataset, CARPK, and introduced a method to generate feasible region proposals, which will be helpful for the drone-based vehicle counting method using the deep learning method. Van Gemert et al. (2014) proposed combining UAVs with automatic object recognition techniques to solve the manual animal surveying problem. Reinecke and Prinsloo (2017) expounded the benefit of drone monitoring on agriculture. A drone can effectively capture detailed pictures and videos to help farmers observe crops remotely and cultivate them better, and scare away the birds (Hill, 2000). Based on all these studies, it can be seen that there is a huge potential for the applications of drones in agriculture. From a livestock perspective, drones are being used to perform headcount and animal monitoring. Cow counting with drones saves time and is more accurate, because an aerial maneuver can provide a clearer multi-view perspective of the scene without cows blocking each other along the line of sight.

In some countries, farmers are already using a live stream drone, but they still count animals in the stream video manually. Al-Thani et al. (2020) assembled and installed a drone by themselves attempting to count and monitor sheep on a farm. They compared offline processing, online pre-processing, and online processing findings and concluded that online processing had the highest accuracy. Real-time processing also helps making just-in-time decisions. Shao et al. (2020) and Rivas et al. (2018) utilized convolutional neural networks on UAV image detection. Hodgson et al. (2018) considered drones as suitable devices to detect and count objects. Despite the above studies, counting the number of cows using drones is still a challenging problem when the herd is too large scattered in open pastures. To address this problem, we designed a machine learning based framework that can automatically count the number of cows via a live stream drone. Based on the evaluation in our pre-trained model feasibility study, we found that the Mask R-CNN (He et al., 2017) model had the best performance. Note that head counting is not sufficient for cattle welfare. It is important to localize individual cows and understand if a cow displays healthy postures and is socially accommodated in the herd. To check the location of each cow, our framework plots its relative position in a 3D canvas, which can be navigated in a 360° viewpoint so that the viewer can monitor the health state of the entire herd.

2. METHOD AND IMPLEMENTATION

![Figure 1. Proposed system processing pipeline](image-url)

In this section, we will introduce our proposed cow counting and 3D visualization system. Figure 1 illustrates the processing pipeline. We first capture videos of the pasture using a drone, and then detect and count the cows. The pipeline is composed of different stages as explained in the subsections below.

2.1 Data Fetching

Due to the constraint imposed by the machine learning model, only one image is allowed for every input. Therefore, we had to divide the captured video into a sequence of frames, detect cows and record their positions in each frame. In other words, for each frame of the video, we generated a headcount and a list of cow positions.
2.2 Cow Detection

We detected cows using two pre-trained models and compared their accuracy performance. The first one was YOLO-v3 (Redmon & Farhadi, 2018). We installed MXNet (Chen et al., 2015) of the GluonCV Toolkit package. Following the model’s input structure, we prepared each image into three grids, 8*8, 16*16, and 32*32. We found that YOLO (You Only Look Once) does not have the highest accuracy, but it is a good compromise between accuracy and speed. YOLO-v3 was an enhanced version, based on a combination of YOLO-v1 (Redmon et al., 2016) and YOLO-v2 (Redmon & Farhadi, 2017). It improves the detection accuracy, while maintaining the speed advantage of other YOLO models, especially when detecting small objects. The YOLO-v3 model divides the image into multiple regions of different sizes. It then classifies each object bounding box based on the calculated probability. YOLO-v3 implements the Darknet-53 model structure, which has 53 convolution layers. Each convolution layer is followed by a batch normalization layer and a leaky ReLU layer as the activation function. The detected result is not ideal, as the speed of the model is too slow, requiring around 10 seconds per frame, which cannot support a real-time application. The accuracy is also not sufficiently high, especially when the cows in the image stack on each other, causing double counting.

For the above reasons, we tested another pre-trained Mask R-CNN model. We installed the Detection 2 (Wu et al., 2019) package, which is made available by Facebook and import R50-FPN-3x. This model can only be tested on the Google Colab on Mac or Linux systems. Mask R-CNN performs an instance segmentation, which not only needs to localize the objects correctly, but also needs to segment the detected objects accurately. Therefore, instance segmentation combines both object detection and semantic segmentation. Mask R-CNN is an improved Faster R-CNN (FCN) (Ren et al., 2015). It semantically segments each proposal candidate box. There are three steps in the Instance Segmentation Mask R-CNN framework: (1) target detection – draw a bounding box around the target object to mark what is detected; (2) target classification – for each target, find the corresponding class to determine whether it is a person, a car, or other categories; and (3) pixel-level target segmentation. For each target, it is necessary to separate the foreground and background. The output of the model is the bounding box position, classified label, and label probability. We count the number of the bounding boxes in each frame as the number of cows.

2.3 Cow’s 3D Visualization

We compared two publicly available free software tools: Blender and Matplotlib 3D, to visualize cows and their movements in a 3D canvas. Blender cannot update the 3D scene frame by frame; it only shows the final scene at the end. Therefore, we selected Matplotlib 3D. Based on the position of each bounding box in the video frame, we placed a cow model in the 3D canvas. We used the x-coordinate and y-coordinate of the bounding box in the image to represent the horizontal position of the cow, and the distance between the cow and the camera in the 3D canvas. The runtime of Matplotlib 3D visualization is around 2 frames per second.

3. EXPERIMENTAL RESULTS

The runtime and accuracy of the two cow detection models are shown in Table 1. Compared with Mask R-CNN, YOLO-v3 is too slow. It cannot meet the requirements of real-time applications. The final accuracy of the YOLOv3 model is 14/15 after scanning every video frame, and 15/15 is the accuracy of Mask R-CNN. Since both the herd and the drone are moving, detection error can easily happen when the scanning device changes direction, which can blur the entire frame. As shown in Figure 2, Mask R-CNN can clearly detect every cow in the herd, even if only a small part of the cow is exposed. However, YOLO-v3’s result is not accurate. YOLO-v3 cannot distinguish each cow in a group when they are of the same color crowding together. Therefore, we used the detection result of Mask R-CNN for 3D visualization.

<table>
<thead>
<tr>
<th>Method</th>
<th>Runtime</th>
<th>Detected Value/True Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>YOLO-v3</td>
<td>10s per frame</td>
<td>14/15</td>
</tr>
<tr>
<td>Mask R-CNN</td>
<td>2.18s per frame</td>
<td>15/15</td>
</tr>
</tbody>
</table>
Figure 2. An example comparing the detection results of frame50 using (a) YOLO-v3 and (b) Mask R-CNN

Figure 3. 3D Visualization of frame50 using Matplotlib 3D

Figure 3 is an example of the 3D visualization output. The current video frame and its detection result are shown at the top, and the number detected is shown at the upper left corner of the frame. The frame number is displayed below the frame. The bottom half of Figure 3 shows the corresponding 3D visualization of the herd in the current frame. Each detected cow is simulated as a 3D point and their positions are updated based on the current frame. Figure 3 (a) and (b) represent the different viewing angles of the same frame. The viewer can switch the viewing perspective to monitor herds and control the flight route of the drone accordingly.

Our 3D visualization framework combined with the UAV, i.e., drone, technology can monitor herds scattering across open fields in real-time. The cow counting result and the 3D visualization show that our system is more effective and accurate than a manual counting process.

4. CONCLUSION

The ability to accurately and quickly perform cattle headcount and monitor their movements in open fields is a time-saver for busy ranchers with big herds. The process has important financial implications. In this paper, we review state-of-the-art methods and develop a real-time cow counting and 3D visualization framework. The processing pipeline starts from taking videos from the drone, localizing the targets, counting the number of cows using object detection, to 3D visualization of the relative position of each cow. We analyzed and compared different object detection models and selected the best performed one to achieve our goal. Monitoring and counting cattle using drones is a challenging problem because it involves a dynamic scene and a moving scanning device. There are many factors that can impact accurate cattle detection and counting, e.g., lack of contrast, cow movements, large cow clusters, and hidden calves. In future work, we will collect more videos for model training. The speed of 3D rendering can also be improved by using more powerful GPU.
REFERENCES


