26th International Congress on Modelling and Simulation, Adelaide, SA, Australia, 30 Nov–4 Dec 2025
mssanz.org.au/modsim2025

A CUDA-optimised key frame extraction for biometric labelling in dairy cattle identification

Md Farhad Hasan [a], **Elizabeth M Morse-McNabb**[b,c]

[a] Agriculture Victoria Research, La Trobe University AgriBio, Victoria, Australia, [b] Agriculture Victoria Research, Ellinbank Smart Farm, 1301 Hazeldean Road, Ellinbank, Victoria, Australia, [c] School of Applied Systems Biology, La Trobe University, Bendigo, Victoria, Australia
Email: farhad.hasan@agriculture.vic.gov.au

Abstract: We implemented a CUDA-accelerated frame extraction process to convert video files into image data to develop an artificial intelligence model for biometric identification and tracking of dairy cattle. To serve this purpose, we used CUDA Python with PyTorch to develop a model that converts high-resolution videos into key frames. The pipeline leverages GPU parallel computing on an NVIDIA Tesla V100 (32 GB) in combination with 36-core CPU parallelism. Cow muzzles are our region of interest. The muzzle consists of the central region of the upper lip and the extension of this upwards between the nostrils. The muzzle prints of dairy cattle serve as unique biometric features analogous to human fingerprints. The extracted frames are assessed and stored for annotation for future modelling workflow for biometric identification.

Keywords: *CUDA Python, key frame extraction, image dataset*

1. INTRODUCTION

As livestock systems continue to implement more technology-driven practices, biometric identification and tracking have become important to advancing, data collection, analysis, modelling and simulation capabilities, because biometrics can be assessed without additional wearable technology. Within this context, the muzzle region of dairy animals has been identified as a unique biometric feature, comparable to fingerprints in humans, and therefore has been used for cattle tracking and identification in smart agricultural systems (Mahato and Neethirajan, 2024). However, identifying and tracking individual dairy animals using the muzzle feature requires a dataset of high-quality images to train and validate an artificial intelligence (AI) model.

Current practices for image collection typically involve manual, individual animal photography. Despite its reliability, this approach is labour-intensive, tedious, and may cause distress to animals through repeated and close-contact handling. As part of our broader efforts to improve data readiness for biometric identification modelling, we have implemented an alternative approach that involves continuous video recording as dairy animals exit through a race in a sequence, one after another. This method reduces manual interference while facilitating large-scale, passive data acquisition under standard herd movement. However, the obtained video recordings generally produce high-volume, unstructured datasets that require further processing to extract the relevant frames. To address this, we have developed a CUDA-accelerated key frame extraction pipeline, implemented in CUDA Python, that rapidly converts raw videos into images of individual dairy animals with visible muzzles, suitable for labelling. This data pipeline is designed to support the development of biometric datasets for modelling, focusing on real-time cattle tracking and identification.

We aimed to improve the practicality and efficiency of dataset curation based on the video footage obtained from a commercial dairy farm. The method described here is one essential component of the future cattle biometric identification model. Preliminary results from the key frame extraction, including processing time, dataset curation, integration potential, and finally, the model accuracy will be briefly discussed.

2. METHODOLOGY AND RESULTS

The data collection started at Ellinbank Smart Farm in November 2024 and is ongoing. All recording was done with two GoPro HERO12 Black (GoPro Inc., San Mateo, CA) action cameras in 1080p resolution at 30 frames per second. The HyperSmooth 6.0 feature of the GoPro Hero 12 provided stabilised video quality despite the animal movements. The cameras were positioned discreetly at the exit point of two cattle races on the smart farm. Figure 1 represents the workflow diagram of this study.

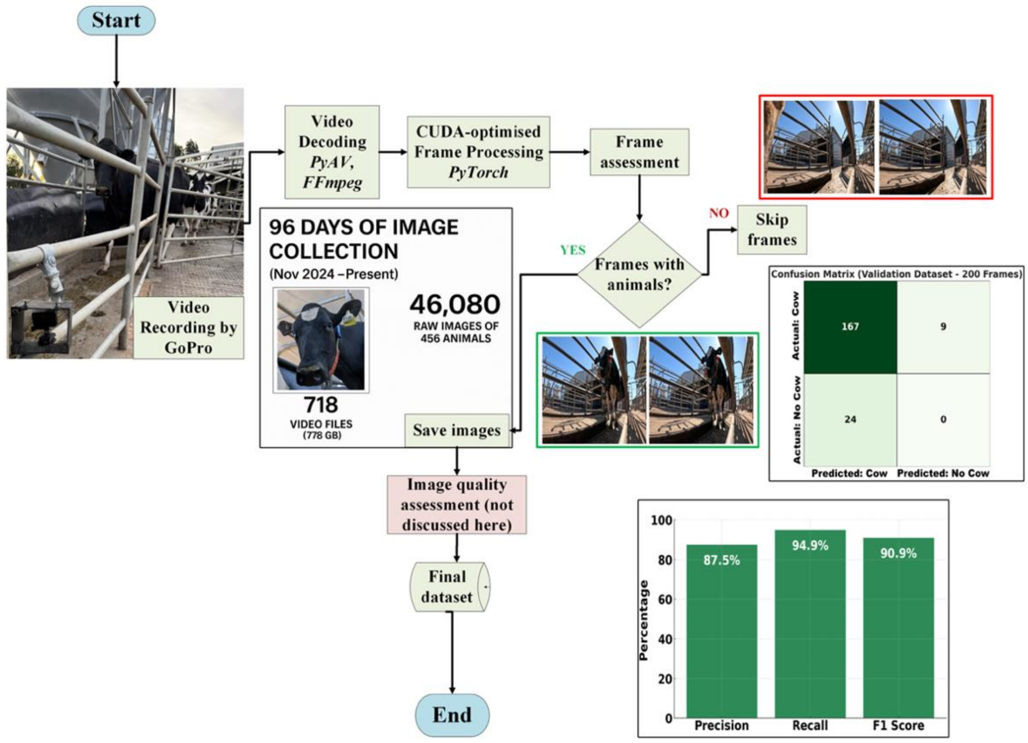

Figure 1 **The workflow diagram and outcome of present study**

The recorded video files were subject to key frame extraction by PyTorch, which is a tensor-based deep learning framework. The computational task was performed using an NVIDIA Tesla V100 (32GB). The CPU parallelism was established in the pipeline using *'concurrent.features'* across 36 physical cores. The video decoding was established by PyAV, a lightweight Python interface to FFmpeg that allowed efficient access to frames without loading the entire video footage into memory. Each video frame was decoded into a NumPy array and then converted to a PyTorch tensor. PyTorch tensor was then used to transfer information to the GPU memory for the frame evaluation based on the presence of cows. To enable the model to identify dairy animals without manual labelling, we used a pre-trained YOLOv5, and trained on the Common Objects in COntext (COCO) datasets (20: u'cow'). The purpose was to identify key frames where the animals were visible rather than identifying individual animals accurately. This approach enabled the exclusion of frames lacking animal presence, thereby ensuring that subsequent processing and annotation efforts were focused exclusively on frames containing cattle. The images were saved based on the three- or four-digit ear tag numbers to identify animals, and an image quality assessment method to assess the suitability of those frames for labelling. The model processed a 60-minute 1080p video in under 3 minutes, including decoding, filtering, and storing the extracted key frames. Validation against a manually annotated frame-by-frame dataset revealed that the model attained a precision of 87.5%, a recall of 94.9%, and an F1 score of 90.9%. The validation dataset was derived from the same commercial dairy environment as the main dataset, capturing variability in lighting, cow positioning, and background activity.

3. CONCLUSIONS

We have presented a CUDA-accelerated video frame extraction pipeline that is robust and efficient. It improves the efficiency of dataset curation for future modelling workflows. Our method uses footage from an active commercial dairy farm with reduced processing time that enables us to form a dataset of muzzle visibility by discarding irrelevant frames. We aim to continue to assess the frames for future AI model development. In the next part of our study, we aim to automatically annotate muzzle images to train a dairy cattle identification model.

ACKNOWLEDGEMENTS

The work is funded by the Victorian Dairy Innovation Agreement, a joint venture between Dairy Australia, Gardiner Dairy Foundation, and Agriculture Victoria (Victoria State Government).

REFERENCES

Mahato, S, and Neethirajan, S (2024). Integrating artificial intelligence in dairy farm management− biometric facial recognition for cows. Information Processing in Agriculture, https://doi.org/10.1016/j.inpa.2024.10.001

