Augmenting Cattle Tracking Efficiency Through Monocular Depth Estimation

Lewis T. Dickson*¶, Christopher Davison*, Craig Michie*, Ewan McRobert*, Robert Atkinson*, Ivan Andonovic*, Holly Ferguson†, Richard Dewhurst†, Roger Briddock‡, Mark Brooking‡, Dejan Pavlovic§, Oskar Marko§, Vladimir Crnojević§, Christos Tachtatzis*

*Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, G1 1RD, UK

†Scotland's Rural College (SRUC), King's Buildings, West Mains Road, Edinburgh, EH9 3JG, UK

‡First Milk, 1 George Square, Glasgow, G2 1AL, UK

§BioSense Institute, Dr Zorana Dindica Street 1, Novi Sad 21000, Serbia

¶Corresponding Author; Email: lewis.dickson@strath.ac.uk

Abstract—We present a method for 3D cattle tracking and inter-camera pose transformation using depth information from monocular depth estimation with deep networks. Camera-based animal monitoring offers a minimally invasive and easily adaptable solution for tracking and welfare monitoring, relying solely on commercial RGB camera systems. However, environmental factors and inter-animal occlusion often hinder tracking efficacy and consistency. To address these challenges, we developed a pipeline to extract 3D point cloud data of individual cows in a straw-bedded calving yard environment, generating quasi-3D bounding boxes $(x, y, z, \text{height}, \text{width}, \theta)$, where θ is the polar angle. We then estimate the camera system extrinsic parameters by minimising the rotation, translation, and scale discrepancies between the apparent motion of animals across different frames of reference. This approach demonstrates a strong agreement between the 3D centroids of tracked animals in motion. Our work advances the development of algorithmic occlusion handling and object handover techniques in multi-camera systems, particularly pertinent to the high-occlusion, low-locomotion scenario of animals within barn environments.

Index Terms—Precision Farming, Monocular Depth Estimation, Tracking, Camera Calibration

I. Introduction

Increasingly automated farms require more effective animal monitoring [1] as the ratio of farmers to cattle decreases. Precision livestock farming of dairy cattle has evolved significantly in recent decades and the use of automated monitoring tools including, for example, accelerometer collars for behavioural recognition [2], tail position monitors to identify the onset of parturition [3], and bolus sensors for rumination duration and frequency are commonplace [4]. However, these devices can affect the safety, welfare, and comfort [5] of livestock and rely on human intervention for implementation and replacement.

Monitoring stock without the need for physically attaching sensors is desirable from a welfare perspective and to minimise operational costs. With the increasing availability of neural network-based image analysis techniques, multiple object tracking has seen a rapid increase in interest [6]. Vision-based approaches for animal tracking, such as using single [7], [8] and distributed [9] camera systems within barn environments have already been implemented.

However, vision-based tracking faces significant challenges in becoming a reliable diagnostic tool for animal welfare. Accurately identifying and tracking individual animals in crowded and obstructed environments—where occlusion due to environmental and inter-animal factors is common—is difficult to achieve [10]. One proposed solution involves utilising depth information alongside RGB-based tracking to simplify the process of handling object-background segmentation and occlusion challenges [11]. Expanding from 2D to 3D tracking improves individual localisation that can be used to optimise barn design and access to resources, quantify social interaction and potentially quantify known positive welfare traits such as synchrony within herds [12]. It also eases the challenge of identifying affiliative and agonistic behaviours in cattle [13] such as mounting, displacement, or allogrooming [14], important indicators of oestrus [15], or hierarchy and social bonds within groups [13], [16].

Recent advances, have employed depth information alone [17], [18] or in conjunction with RGB image data to augment cattle identification [19], [20] and postural or characteristic traits [21]–[24]. Production of the RGB-Depth (RGB-D) maps in the above examples was achieved via amplitude-modulated continuous-wave time-of-flight principles [25] or stereoscopic scene reconstruction. Consumer-grade RGB-D sensors (e.g. the Microsoft Kinect sensor) are, however, sensitive to lighting conditions [26] and unsuitable for large barn environments as their depth measurement range is limited to approximately 5 metres [27]. Stereoscopic approaches [28], are similarly constrained resulting in sparse and incomplete 3D scene reconstruction [22].

In this work, we employ Monocular Depth Estima-

The work was conducted under the auspices of the UKRI Digital Dairy Value-Chain for South-West Scotland and Cumbria (Strength in Places Fund) award application number is 99890. Ethical approval for the experiment was not required because the study did not affect the animals being observed.

tion (MDE) through deep learning [29] to generate 3D representations of the scene. This approach aims to enhance tracking efficiency and enable object handover by accurately capturing the spatial and angular distribution of scene objects. A deep learning-based approach was selected over traditional MDE methods, such as structure-from-motion, which are less well suited to bovine tracking due to the prolonged periods of cattle inactivity during rumination, and their slow movement within barn environments.

Assessing the feasibility of an end-to-end depth-aware tracking solution, we have investigated currently available methods where we have prioritised efficacy over computational time. Namely, we apply Depth Anything [30] for estimating depth maps from single RGB camera images and develop a pipeline for extraction of 3D point cloud motion over time.

The paper is organised as follows. Section II presents the experimental animal monitoring setup within a straw-bedded calving yard environment. In Section III we demonstrate the production of RGB-D information from video feeds using monocular depth estimation methods and calculating the plane-of-best-fit for the point cloud associated with a given bounding box. Section IV demonstrates the time-dependent point clouds of a tracked animal under motion from two opposing camera views. In Section V the location and angular distribution of the planes-of-best-fit are used to track cattle in 3D with two cameras exploiting the apparent paths to externally calibrate the rotated and translated pose of the two-camera system.

II. Experimental Design

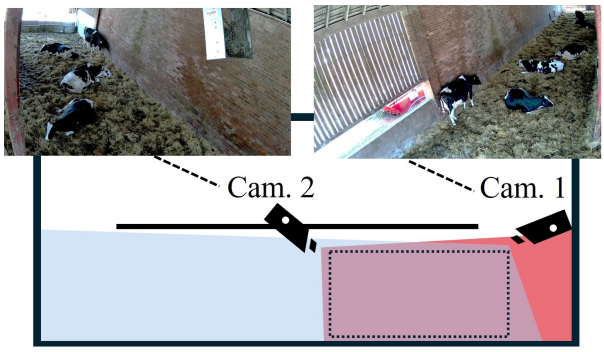

Fig. 1. Positioning and field of view for two cameras within the straw-bedded calving yard. Shaded regions in blue and red show the coverage of Cam. 1 and Cam. 2, respectively. The dashed region formed by the overlapping view is used for animal tracking within this study. The black horizontal line represents an internal railing. Cam. 2 is mounted on a pillar of this railing.

Video monitoring samples were collected at the Dairy Research and Innovation Centre at Crichton Royal Farm using consumer-grade Reolink Duo cameras situated within the straw-bedded calving yard. The yard is a 12m by 35m rectangular building with a fence running along the mid-line of the yard with access to both sides allowing free roaming. Movable fences within the yard facilitate sectioning a subset of cows for handling or to allow an individual to be isolated from the herd during calving. The yard is instrumented with 12 dual-lens cameras, giving 24 video feeds. Of the 24 video feeds available within the yard, the two camera systems used in this work are implemented as shown in Figure 1. The manufacturer-specified focal lengths were confirmed offline using ChArUco board calibrations and were within $< 2\%$ error in vertical and horizontal directions and so are assumed equal in the following analysis.

III. RGB-D Data Production Pipeline

Video streams are converted into RGB-D videos using the method illustrated in Figure 2. Classification and production of 2D bounding boxes are extracted using the standard YOLOv8 model [31] applied frame-wise to the video sample with BoT-SORT [32] enabled to perform the tracking; an example output is shown in Figure 2 B). In parallel, each frame is supplied to the Depth Anything [30] model using their ViT-L encoder and metric indoor model weights which produced the depth map at each pixel coordinate, shown in C). To increase the resolution of the depth map, we used the method presented in Boost Your Own Depth [33] with fixed 4×4 and 8×8 pixel filter sizes which produced the depth map in E). Following the production of the bounding boxes, the animal silhouettes were segmented, as shown in D), using the bounding boxes as anchor positions with the Segment Anything model [34]. Depth values were extracted from within these masks which created a 3D point cloud corresponding to the 3D position of the animal within the barn. A plane-of-best-fit was then computed for each point cloud to extract the centroid (via the 3D position of the point cloud) and angle (via the plane normal) of each segmented animal. The resulting point clouds and their associated planes are displayed in F). This process was repeated for each frame and camera view leading to the production of the 3D position of the animal in time as illustrated in Figure 3. An example still image of the 3D rendering of the RGB-D is shown in Figure 2 G) illustrating a false view of the scene with the cow of interest standing in the foreground.

Scale ambiguity in MDE is a recognised challenge, where depth can be determined up to an unknown scale factor. While this analysis operates on arbitrary depth scales, metric measurements can be achieved using ChArUco marker systems, rescaling known distances in the barn environment, or employing spatially calibrated cameras with estimated object sizes like cows, gates, or pillar distances. Refining metric depth calibration using these methods is the focus of ongoing research.

IV. 3D Point Cloud Mapping

The RGB-D data supports the extraction of additional information on the motion and location of the cattle at

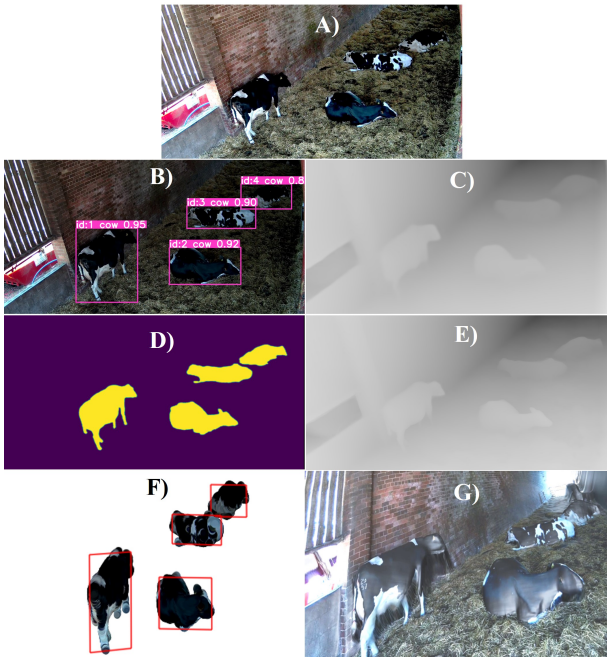

Fig. 2. Illustration of processing RGB to RGB-D data for each video frame. A) RGB frame extracted from a video feed. B) YOLOv8 classification, bounding boxes, and confidence score for the RGB frame A. C) Depth Anything MDE predicted depth map from A. D) Segment Anything bounding box-anchored segmentation using A and B. E) Resolution-boosted depth map from C using tiling and low-resolution depth averaging method from Boost Your Own Depth. F) RGB-D 3D point clouds and associated plane-of-best-fit (red) using the masks from D applied to A for RGB and E for depth. G) 3D rendering of the dense RGB-D scene.

rest or when active within the barn. Figure 3 demonstrates the extraction of the point clouds over time for a moving cow viewed by two approximately anti-parallel cameras. Figure 3 A) illustrates the animal motion from the front corner of the barn to the central region. Note that the blurred regions in the frames of Cam. 1 and Cam. 2 (on the right and left sides, respectively) are caused by structural occlusions to the camera system, where depth information is unknown. This blurring is rendered due to the false viewing angle used to visualise the RGB-D frames.

Figure 3 B I) and II) demonstrate the motion of the cow every 10^{th} frame of the video sample for Cam. 1 and Cam. 2, respectively, with the time progression indicated by the transition from light to dark colours. Given the anti-parallel camera views, the cow appears to move away from and towards Cam. 1 and Cam. 2, respectively, with time. Fitting a plane-of-best-fit to each masked point cloud using the method described in Figure 2 F), and illustrated only for the moving animal within the scene, allows us to calculate a more accurate centroid of the animal by extracting the centre and angle of the animal under motion. With this approach, we can expand the 2D bounding box to quasi-3D by including the polar (θ) angular component extracted from fitting a plane to the point cloud in addition to the centroid (x, y, z), plus the width and height of the 2D bounding box. The point cloud time sequence (from lighter to darker) illustrates that the proposed MDE method is sufficiently sensitive to extract the position of the animal as it walks away from (Figure 3 B I)) and towards (B II)) the cameras and the angular motion of the cow agree with the motion seen in panel A).

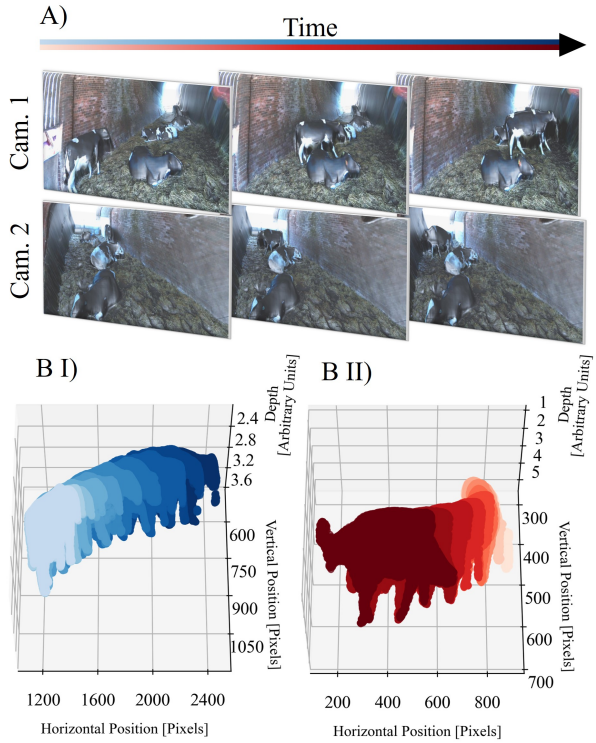

Fig. 3. Frame-by-frame extraction of point cloud for 3D cow position for multiple camera views. A) Example rendered RGB-D frames produced using the proposed cow moving within the barn for two opposing camera views. B I) and II) 3D point cloud sequence (every 10 frames) for the same cow for Cam. 1 and Cam. 2, respectively, where time progression is indicated by colour changes from lighter to darker.

V. Point Cloud Centroid Tracking and Inter-Camera Calibration

Improving tracked object handover in 3D necessitates measuring the volumetric intersection-over-union for the quasi-3D bounding boxes. A transformation between the camera viewpoints is consequently required to create the bounding box overlap in a shared coordinate system. Calculating the position and rotation of the camera viewpoint within a global coordinate system, known as the extrinsic parameters, typically requires spatial calibration markers or additional object viewing angles. Implementing these methods is challenging for opposing camera views, such as in this example.

To approximate the rotation, translation, and scaling between the two cameras, we use the frame-wise 3D centroid of the segmented point clouds as a minimisation target to align the apparent paths from different viewpoints.

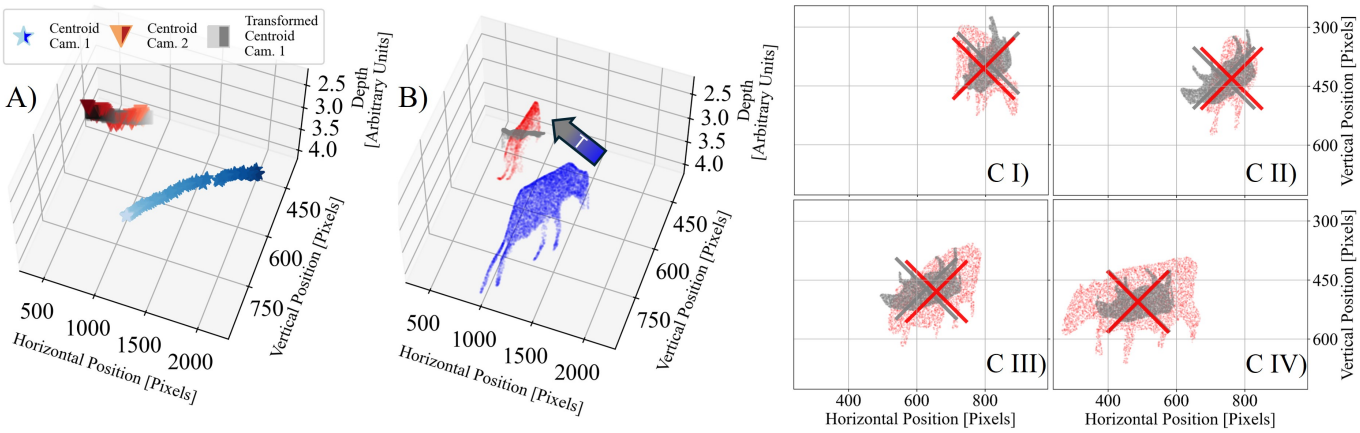

Fig. 4. Extraction of the 3D tracking centroid paths and its application to camera extrinsic extraction and object overlay. A) Frame-wise 3D point cloud centroid motion from the view of Cam. 1 (blue stars), Cam. 2 (red triangles), and the Kabsch-transformed position vector of Cam. 1 (grey squares). Lighter colours indicate earlier frames. B) Kabsch transform applied to point cloud where the arrow 'T' indicates the transform direction (from blue to grey.) C) Depth-flattened view of the transformed Cam. 1 (grey) and reference Cam. 2 (red) point clouds for every 20$^{\text{th}}$ frame, chronologically ordered by $t_{\text{CI}} < t_{\text{CII}} < t_{\text{CIII}} < t_{\text{CIV}}$. Crosses correspond to the horizontal and vertical components of the 3D centroid of each point cloud for their corresponding colour.

This is computed using the Kabsch algorithm [35] with paired 3D points, which calculates the required translation from the average centroid of all points in each point set, and the rotation from the singular value decomposition of the covariance matrix after translation. Taking inspiration from Procrustes analysis [36], we have used a modified form of the Kabsch algorithm that incorporates uniform scaling to model the different object distances from the cameras, but without unit scale normalisation.

The Kabsch transformation is then applied to shift the path produced from the view of Cam. 1 into the frame of reference for Cam. 2. Figure 4 A) shows 3D point cloud centroid motion from the views of Cam. 1 and 2 in addition to the Kabsch-transformed position vector of Cam. 1 to Cam. 2. Note the axes denote the same coordinate system as Figure 3 B. In these trajectories, lighter colours represent earlier frames in the sequence. After applying the Kabsch transformation, the transformed and reference trajectories for Cam. 1 and Cam. 2, respectively, are well aligned in terms of path position, length, angle, and direction as seen in Figure 4 A). The accuracy of this transform could be improved by averaging the estimated rotation and translation matrices calculated for multiple animal paths, and by including skew, both of which are the focus of future work.

This transformation is applied to the full point cloud of the tracked cow from Cam. 1 (blue cow) leading to the spatial overlap of the transformed point cloud from Cam. 1 (grey) and the reference view of Cam. 2 (red cow) in Figure 4 B). Note that the transform produces an axially inverted image to account for the apparent flipped motion of the animal between the approximately anti-parallel camera views. Owing to the limitations of illustrating overlapping 3D views in print, depth-flattened views of the transformed (grey) and reference point cloud of Cam. 2 (red) for every 20$^{\text{th}}$ frame are shown in Figure 4 C). The good agreement between the centroids of the transformed and reference 3D point clouds projected back into 2D indicates that this method can provide cross-identification of the bounding boxes produced by the tracker for multiple camera views. The successful overlap of the point clouds across all frames indicates that the 3D centroids used to calculate the Kabsch transform adequately describe the animal motion. Additionally, this could be directly applied to volumetric intersection-over-union matching between multiple camera systems and improved handover during occlusions by matching the expected motion of the animal between one view and the other.

VI. Conclusions

This study presents a proposed pipeline using monocular depth estimation via deep networks to generate RGB-D data from consumer-grade RGB camera systems in a straw-bedded calving yard environment. RGB-D data was used to calculate frame-wise 3D centroid and quasi-3D bounding box information for tracked animals. An image transform procedure was proposed to extract approximate camera extrinsic parameters for future volumetric intersection-over-union cross-referencing detection. Depth information can provide monitoring of inter-animal behaviours expanding on the repertoire of welfare indicators that can be tracked. In addition, the improved reliability of individual animal tracking will lead to higher precision in animal welfare monitoring by reducing animal mismatch and intermittent tracking. Work is underway to implement this method within a depth-aware tracking system for multi-camera systems to improve the stability, and object handover, of tracking processes.

References

[1] H. Barkema, M. Von Keyserlingk, J. Kastelic, T. Lam, C. Luby, J.-P. Roy, S. LeBlanc, G. Keefe, and D. Kelton, "Invited review: Changes in the dairy industry affecting dairy cattle health and welfare," Journal of Dairy Science, vol. 98, no. 11, pp. 7426–7445, Nov. 2015.

[2] D. Pavlovic, C. Davison, A. Hamilton, O. Marko, R. Atkinson, C. Michie, V. Crnojević, I. Andonovic, X. Bellekens, and C. Tachtatzis, "Classification of Cattle Behaviours Using Neck-Mounted Accelerometer-Equipped Collars and Convolutional Neural Networks," Sensors, vol. 21, no. 12, p. 4050, Jun. 2021.

[3] S. Higaki, H. Okada, C. Suzuki, R. Sakurai, T. Suda, and K. Yoshioka, "Estrus detection in tie-stall housed cows through supervised machine learning using a multimodal tail-attached device," Computers and Electronics in Agriculture, vol. 191, p. 106513, Dec. 2021.

[4] A. Hamilton, C. Davison, C. Tachtatzis, I. Andonovic, C. Michie, H. Ferguson, L. Somerville, and N. Jonsson, "Identification of the Rumination in Cattle Using Support Vector Machines with Motion-Sensitive Bolus Sensors," Sensors, vol. 19, no. 5, p. 1165, Mar. 2019.

[5] A. Herlin, E. Brunberg, J. Hultgren, N. Högberg, A. Rydberg, and A. Skarin, "Animal Welfare Implications of Digital Tools for Monitoring and Management of Cattle and Sheep on Pasture," Animals, vol. 11, no. 3, p. 829, Mar. 2021.

[6] Y. Park, L. M. Dang, S. Lee, D. Han, and H. Moon, "Multiple Object Tracking in Deep Learning Approaches: A Survey," Electronics, vol. 10, no. 19, p. 2406, Oct. 2021.

[7] C. C. Mar, T. T. Zin, I. Kobayashi, and Y. Horii, "A Hybrid Approach: Image Processing Techniques and Deep Learning Method for Cow Detection and Tracking System," in 2022 IEEE 4th Global Conference on Life Sciences and Technologies (LifeTech). Osaka, Japan: IEEE, Mar. 2022, pp. 566–567.

[8] C. C. Mar, T. T. Zin, P. Tin, K. Honkawa, I. Kobayashi, and Y. Horii, "Cow detection and tracking system utilizing multi-feature tracking algorithm," Scientific Reports, vol. 13, no. 1, p. 17423, Oct. 2023.

[9] S. Han, A. Fuentes, S. Yoon, Y. Jeong, H. Kim, and D. Sun Park, "Deep learning-based multi-cattle tracking in crowded livestock farming using video," Computers and Electronics in Agriculture, vol. 212, p. 108044, Sep. 2023.

[10] W. H. E. Mg, P. Tin, M. Aikawa, I. Kobayashi, Y. Horii, K. Honkawa, and T. T. Zin, "Customized Tracking Algorithm for Robust Cattle Detection and Tracking in Occlusion Environments," Sensors, vol. 24, no. 4, p. 1181, Feb. 2024.

[11] A. Lukezic, U. Kart, J. Kapyla, A. Durmush, J.-K. Kamarainen, J. Matas, and M. Kristan, "CDTB: A Color and Depth Visual Object Tracking Dataset and Benchmark," in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Oct. 2019.

[12] R. J. Kilgour, "In pursuit of "normal": A review of the behaviour of cattle at pasture," Applied Animal Behaviour Science, vol. 138, no. 1-2, pp. 1–11, Apr. 2012.

[13] B. Foris, M. Zebunke, J. Langbein, and N. Melzer, "Comprehensive analysis of affiliative and agonistic social networks in lactating dairy cattle groups," Applied Animal Behaviour Science, vol. 210, pp. 60–67, Jan. 2019.

[14] S. Sato, K. Tarumizu, and K. Hatae, "The influence of social factors on allogrooming in cows," Applied Animal Behaviour Science, vol. 38, no. 3-4, pp. 235–244, Dec. 1993.

[15] S. Reith and S. Hoy, "Review: Behavioral signs of estrus and the potential of fully automated systems for detection of estrus in dairy cattle," animal, vol. 12, no. 2, p. 398–407, 2018.

[16] K. Ren, G. Bernes, M. Hetta, and J. Karlsson, "Tracking and analysing social interactions in dairy cattle with real-time locating system and machine learning," Journal of Systems Architecture, vol. 116, p. 102139, Jun. 2021.

[17] A. Sharma, L. Randewich, W. Andrew, S. Hannuna, N. Campbell, S. Mullan, A. W. Dowsey, M. Smith, M. Hansen, and T. Burghardt, "Universal bovine identification via depth data and deep metric learning," arXiv preprint arXiv:2404.00172, 2024.

[18] C. G. Dang, S. S. Lee, M. Alam, S. M. Lee, M. N. Park, H.-S. Seong, S. Han, H.-P. Nguyen, M. K. Baek, J. G. Lee, and V. T. Pham, "Korean Cattle 3D Reconstruction from Multi-View 3D-Camera System in Real Environment," Sensors, vol. 24, no. 2, p. 427, Jan. 2024.

[19] F. Okura, S. Ikuma, Y. Makihara, D. Muramatsu, K. Nakada, and Y. Yagi, "RGB-D video-based individual identification of dairy cows using gait and texture analyses," Computers and Electronics in Agriculture, vol. 165, p. 104944, Oct. 2019.

[20] W. Andrew, S. Hannuna, N. Campbell, and T. Burghardt, "Automatic individual holstein friesian cattle identification via selective local coat pattern matching in RGB-D imagery," in 2016 IEEE International Conference on Image Processing (ICIP). Phoenix, AZ, USA: IEEE, Sep. 2016, pp. 484–488.

[21] N. Jia, G. Kootstra, P. G. Koerkamp, Z. Shi, and S. Du, "Segmentation of body parts of cows in RGB-depth images based on template matching," Computers and Electronics in Agriculture, vol. 180, p. 105897, Jan. 2021.

[22] J. Lu, H. Guo, A. Du, Y. Su, A. Ruchay, F. Marinello, and A. Pezzuolo, "2-D/3-D fusion-based robust pose normalisation of 3-D livestock from multiple RGB-D cameras," Biosystems Engineering, vol. 223, pp. 129–141, Nov. 2022.

[23] T. Van Hertem, A. Schlageter Tello, S. Viazzi, M. Steensels, C. Bahr, C. E. B. Romanini, K. Lokhorst, E. Maltz, I. Halachmi, and D. Berckmans, "Implementation of an automatic 3D vision monitor for dairy cow locomotion in a commercial farm," Biosystems Engineering, vol. 173, pp. 166–175, Sep. 2018.

[24] A. Ruchay, V. Kober, K. Dorofeev, V. Kolpakov, A. Gladkov, and H. Guo, "Live Weight Prediction of Cattle Based on Deep Regression of RGB-D Images," Agriculture, vol. 12, no. 11, p. 1794, Oct. 2022.

[25] H. Sarbolandi, D. Lefloch, and A. Kolb, "Kinect range sensing: Structured-light versus Time-of-Flight Kinect," Computer Vision and Image Understanding, vol. 139, pp. 1–20, Oct. 2015.

[26] I. C. Condotta, T. M. Brown-Brandl, S. K. Pitla, J. P. Stinn, and K. O. Silva-Miranda, "Evaluation of low-cost depth cameras for agricultural applications," Computers and Electronics in Agriculture, vol. 173, p. 105394, Jun. 2020.

[27] K. Khoshelham and S. O. Elberink, "Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications," Sensors, vol. 12, no. 2, pp. 1437–1454, Feb. 2012.

[28] M. Kytö, M. Nuutinen, and P. Oittinen, "Method for measuring stereo camera depth accuracy based on stereoscopic vision," J. A. Beraldin, G. S. Cheok, M. B. McCarthy, U. Neuschaefer-Rube, A. M. Baskurt, I. E. McDowall, and M. Dolinsky, Eds., San Francisco Airport, California, USA, Jan. 2011, p. 78640I.

[29] C. Zhao, Q. Sun, C. Zhang, Y. Tang, and F. Qian, "Monocular depth estimation based on deep learning: An overview," Science China Technological Sciences, vol. 63, no. 9, pp. 1612–1627, Sep. 2020.

[30] L. Yang, B. Kang, Z. Huang, X. Xu, J. Feng, and H. Zhao, "Depth Anything: Unleashing the Power of Large-Scale Unlabeled Data," arXiv preprint arXiv:2401.10891, 2024.

[31] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLO," https://github.com/ultralytics/ultralytics, Jan. 2023.

[32] N. Aharon, R. Orfaig, and B.-Z. Bobrovsky, "Bot-sort: Robust associations multi-pedestrian tracking," arXiv preprint arXiv:2206.14651, 2022.

[33] S. M. H. Miangoleh, S. Dille, L. Mai, S. Paris, and Y. Aksoy, "Boosting monocular depth estimation models to high-resolution via content-adaptive multi-resolution merging," 2021.

[34] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick, "Segment anything," arXiv preprint arXiv:2304.02643, 2023.

[35] W. Kabsch, "A solution for the best rotation to relate two sets of vectors," Acta Crystallographica Section A, vol. 32, no. 5, pp. 922–923, Sep. 1976.

[36] J. C. Gower, "Generalized procrustes analysis," Psychometrika, vol. 40, no. 1, pp. 33–51, Mar. 1975.