# Evaluating Homography Error for Accurate Multi-Camera Multi-Object Tracking of Dairy Cows

Shunpei Aou[a], Yota Yamamoto[b], Kazuaki Nakamura[c] and Yukinobu Taniguchi[d]
*Department of Information and Computer Technology, Faculty of Engineering, Tokyo University of Science, Tokyo, Japan*

Keywords: Multi-Object Tracking, Image Recognition.

Abstract: In dairy farming, accurate and early detecting of the signs of disease or estrus in dairy cows is essential for improving both health management and production efficiency. It is well-known that diseases or estrus in dairy cows are reflected in their activity levels, and monitoring behaviors such as walking distance and periods of feeding, drinking, and lying down time serves as a means to detect these signs. Therefore, tracking the movement of dairy cows can provide insights into their health condition. In this paper, we propose a tracking method that addresses homography errors, which have been identified as one of the causes of reduced accuracy in the location-based multi-camera multi-object tracking methods previously used for tracking dairy cows. Additionally, we demonstrate the effectiveness of the proposed method through validation experiments conducted in two different barn environments.

## 1 INTRODUCTION

The accurate and early detection of signs of disease and estrus in dairy cows is essential for maintaining their health and improving milk production efficiency. These signs are known to be observed in the amount of movement of dairy cows (Mar et al., 2023), and tracking dairy cows is an important step in herd management. Therefore, tracking the movement of dairy cows is an important step toward monitoring the health of individual cows.

We aim to realize a multi-camera multi-object tracking system that can consistently track the movements of dairy cows in a barn. In order to track the movements of dairy cows moving freely around a large barn, we use multiple cameras installed on the ceiling of the barn and track the cows from the video images. As the cows' appearance in the images strongly depends on camera angle, it is difficult to use existing multi-camera multi-object tracking approaches.

(Aou et al., 2024) proposed a location-based multi-camera dairy cow tracking method by focusing on the fact that adjacent cameras have fields of view that partially overlap. Rather than appearance features, this method uses the overlap degree (Intersection over Union) between bounding boxes in tracking objects across the barn coordinates. Using this method, they achieved tracking in a single barn environment (MOTA 85.6%, IDF1 68.9%). A preliminary experiment revealed that the main cause of poor accuracy is the error in tracking individuals (ID switch) that can occur when a dairy cow crosses the shared field of view of the cameras. As shown in Figure 1, when individual A moves from Cam1 to Cam2, homography error in the bounding boxes yielded by homography increases, and there are cases where one bounding box overlaps another (that of individual B). The homography error increases close to the image edges, so the frequency of ID switches increases around the camera boundaries. (Aou et al., 2024) minimized ID switches by offsetting the distortion from projective transformations through the use of rotated bounding boxes, but homography error, which is determined by the camera installation angle, was not considered. Also, since the accuracy of the bounding boxes was not verified, the impact of the accuracy of the detector on the tracking accuracy could not be elucidated.

This paper proposes a method for multi-camera tracking of dairy cows that considers homography error to improve the tracking accuracy. As will be explained later, homography error is dependent on the

[a] https://orcid.org/0009-0002-4728-3979
[b] https://orcid.org/0000-0002-1679-5050
[c] https://orcid.org/0000-0002-4859-4624
[d] https://orcid.org/0000-0003-3290-1041



Aou, S., Yamamoto, Y., Nakamura, K. and Taniguchi, Y.
Evaluating Homography Error for Accurate Multi-Camera Multi-Object Tracking of Dairy Cows.
DOI: 10.5220/0013160900003912
In *Proceedings of the 20th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2025) - Volume 2: VISAPP*, pages 675-682
ISBN: 978-989-758-728-3; ISSN: 2184-4321

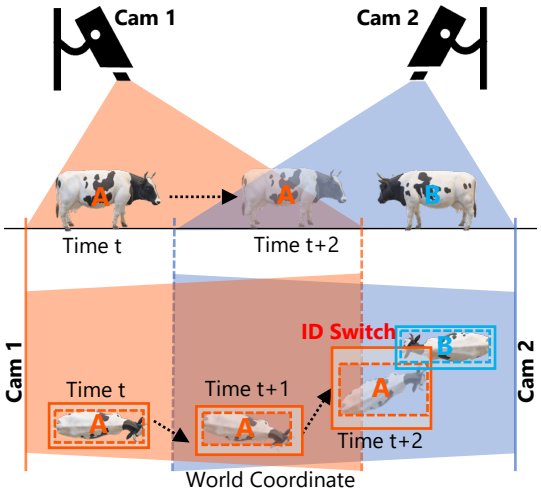



Figure 1: A case of tracking error caused by ID switch. As CowA approaches the edge of Cam1, the bounding box becomes distorted due to the increase in homography error, resulting in ID switch between CowA and CowB.

position and orientation of each camera, so it cannot be dealt with simply by excluding the edges of the camera. In addition, we will verify the generality of the proposed method by confirming its effectiveness in barn environments that differ from those examined in previous research.

The contributions of this paper are as follow: **(i)** a robust tracking method is proposed for multi-camera multi-object tracking that utilizes the homography error and is robust to the camera installation angle, **(ii)** the confirmation of the generality of the proposed method by verifying it in two typical but different barn environments, **(iii)** the comparison of tracking accuracy excluding the influence of detection accuracy by creating ground truth bounding box data in two barn environments.

## 2 RELATED WORKS

**Multi-Camera Multi-Object Tracking.** Object tracking methods can be classified into two categories: those that use the appearance features of the target to be tracked and those that use only location information.

Ristani et al.(Ristani and Tomasi, 2018) proposed multi-camera multi-object tracking using appearance and motion features obtained from a network that has learned the features of objects in advance. Using appearance features makes it possible to perform multi-camera multi-object tracking without location information of cameras. However, it is not suitable for tracking animals. Figure 2 shows the same dairy cows


Figure 2: Appearance features extracted from different angled cameras are significantly different. The spotted patterns of cow ID=5 are completely different between cameras.

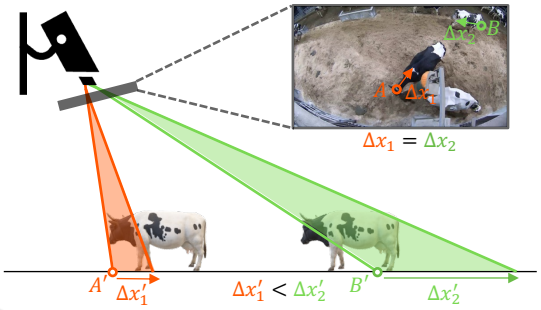



Figure 3: Homography error variation based on the distance between the image center and the cow. $A$ and $B$ are two points in the camera image, $A'$ and $B'$ are their barn coordinates after homography transformation. Given uniform perturbations of equal magnitude, $\Delta x_1 = \Delta x_2$, the error in barn coordinates of $B$, $\Delta x'_2$, will exceed that of $A$, $\Delta x'_1$. This error depends on the angle between the camera, the optical axis, and the floor. If the camera is positioned directly overhead, $\Delta x'_1$ equals $\Delta x'_2$.

as seen by two cameras with a shared field of view. The appearance features of these cows differ greatly between cameras due to the difference in camera orientation.

Location-based multi-camera multi-object tracking methods such as (Bredereck et al., 2012; Eshel and Moses, 2008; Fleuret et al., 2008) have been proposed; they can track using only the location information of the target to be detected. Thus they can also be applied to animal tracking (Akizawa et al., 2022; Aou et al., 2024; Shirke et al., 2021; Yamamoto et al., 2024). (Aou et al., 2024) proposed a location-based multi-camera multi-object tracking method that uses homography for multi-camera tracking of dairy cows. This method (i) detects cows as rotated bounding boxes in all camera images, (ii) integrates them by projecting the rotated bounding boxes obtained from the multi-cameras onto barn coordinates using homography transformation and (iii) tracks cows in the barn coordinates using Simple Online and Real-time Tracking (SORT) (Bewley et al., 2016). By calibrating the correspondence between camera coordinates and barn coordinates in advance, it is possible to project dairy cows into a barn coordinates

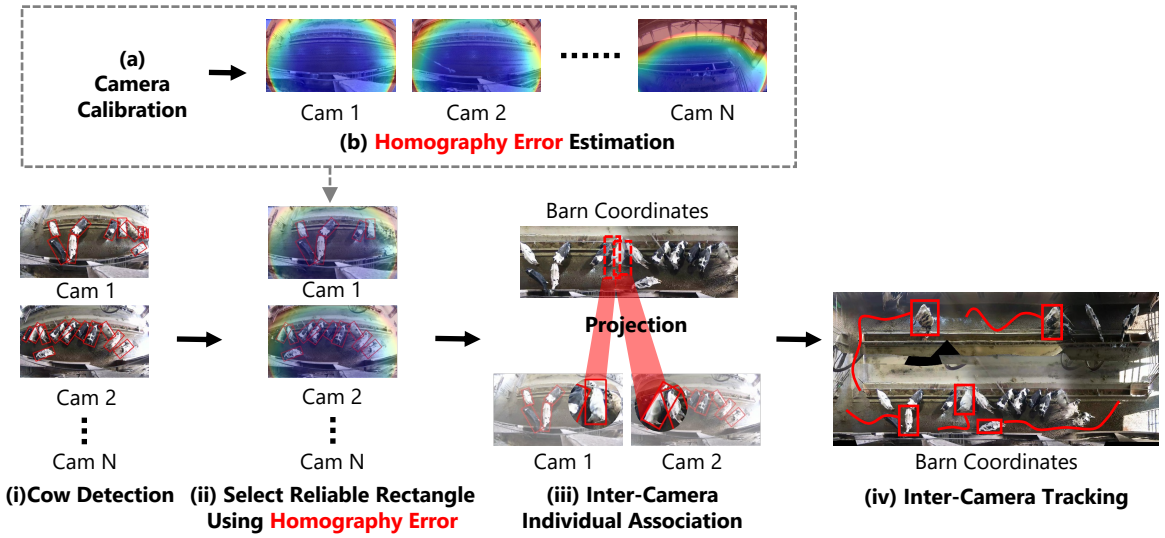



Figure 4: Overview of the proposed multi-camera dairy cow tracking method. In advance, camera calibration and homography error estimation are performed. Online processing consists of 4 steps **(i)** Detecting cows as rotated bounding boxes, **(ii)** Selecting reliable bounding boxes using homography error, **(iii)** Projecting the detected bounding boxes onto a common barn coordinates and integrating the same individual between multiple cameras, **(iv)** Tracking individuals in barn coordinates using the SORT tracking algorithm (Bewley et al., 2016).

and track them throughout the entire barn. However, the method does not account for homography error caused by the distance between the camera and the dairy cow.

**Homography Error in Multi-Camera Multi-Object Tracking.** (Derek et al., 2024) addressed the issue of homography error to improve the accuracy of multi-camera vehicle tracking. Since homography error can occur when the camera shakes due to wind, etc., they reduced homography error beyond that attained by existing image stabilization by performing homography re-estimation. On the other hand, because this study assumes that the cameras are fixed and indoors, the cameras are deemed to have a wide-angle lens, and the distance between the camera and the dairy cow varies depending on the location, so static homography error like the one shown in Figure 3 occurs. The aim is to suppress the homography error that arises from changes in the dairy cow's pose and errors contained in the object detection bounding box, which differs from (Derek et al., 2024) in that it is applied to the tracking method.

# 3 PROPOSED METHOD

We propose a multi-camera tracking method that takes into account homography error. Figure 4 shows an overview of the proposed method. The method consists of two preparation steps and four online steps.

## 3.1 Preparation Steps

**(a) Camera Calibration.** We start with camera calibration to obtain camera parameters and lens distortion coefficients. The extrinsic camera parameters define 3D projective transformation from camera coordinates to the world coordinates.

We assume a pinhole camera model with simple lens distortion:

$$\boldsymbol{p}_d \sim K(S|\boldsymbol{t})P,\ \boldsymbol{p} = f(\boldsymbol{p}_d;\boldsymbol{k}). \quad (1)$$

Here, $\boldsymbol{P} = (X,Y,Z,1)^T$ is the homogeneous coordinates of a 3D point in the world coordinates, and the 3D plane $Z = 0$ is taken as the floor of the barn. $\boldsymbol{p} = (x,y,1)^T$ is the homogeneous coordinates of a 2D point in the camera image, $\boldsymbol{p}_d = (x_d,y_d)^T$ are the homogeneous coordinates after distortion correction, $f$ is the function that corrects the distortion, $K$ and $(S|t)$ are intrinsic and extrinsic camera parameters of the camera, respectively, where $\boldsymbol{t}$ is the translation vector of the camera, and $\boldsymbol{k}$ is the distortion coefficient.

The camera parameters are calculated using Zhang's method (Zhang, 2000) from multiple pairs of 3D world coordinates points $\boldsymbol{Pi} = (X_i,Y_i,Z_i)^T$ and corresponding 2D image points $\boldsymbol{p}_i = (x_i,y_i)^T$. The correspondence between these points is obtained by manually mapping the coordinates in the camera to the world coordinates in advance.

The homography or the planar projection transformation from the camera coordinates to the reference plane $Z = C_h$ parallel to the floor of the barn is calcu-

lated by

$$M = (K(\boldsymbol{s}_1, \boldsymbol{s}_2, t + C_h \boldsymbol{s}_3))^{-1}. \quad (2)$$

In this paper, the 2D coordinates $(X,Y)$ on the reference plane is referred to as 2D barn coordinates, and $C_h$ is defined as the height from the floor to the top of the cow's back (Yamamoto et al., 2024).

**(b) Homography Error Estimation.** If the undistorted camera coordinates $\boldsymbol{p}_d = (x_d, y_d)$ and the homography matrix

$$M = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{pmatrix}, \quad (3)$$

the 2D barn coordinates $\boldsymbol{P}_w = (X_w, Y_w)$ can be obtained by

$$X_w = \frac{m_{11}x_d + m_{12}y_d + m_{13}}{m_{31}x_d + m_{32}y_d + m_{33}}, \quad (4)$$

$$Y_w = \frac{m_{21}x_d + m_{22}y_d + m_{23}}{m_{31}x_d + m_{32}y_d + m_{33}}. \quad (5)$$

Next, partial differentiation of $X_w$ with respect to image coordinates $(x_d, y_d)$ are:

$$\frac{\partial X_w}{\partial x_d} = \frac{m_{11} - m_{31}x_d}{m_{31}x_d + m_{32}y_d + m_{33}}, \quad (6)$$

$$\frac{\partial X_w}{\partial y_d} = \frac{m_{12} - m_{32}y_d}{m_{31}x_d + m_{32}y_d + m_{33}}. \quad (7)$$

Similarly, the partial derivative of $Y_w$ can also be obtained. Using these partial derivatives, the distance from the pixel to the adjacent pixel can be obtained. If the position of pixel $\boldsymbol{p}_d$ is perturbed by $\Delta\boldsymbol{p}_d = (\Delta x_d, \Delta y_d)^T$, the displacement in the 2D barn coordinates system $\Delta\boldsymbol{P}_w = (\Delta X_w, \Delta Y_w)^T$ can be approximated as:

$$\Delta\boldsymbol{P}_w \approx \frac{\partial(X_w, Y_w)}{\partial(x_d, y_d)} \Delta\boldsymbol{p}_d, \quad (8)$$

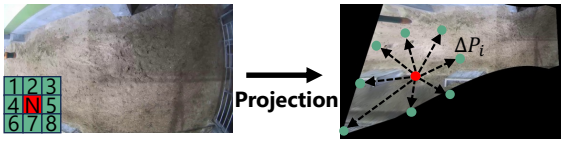
where $\partial(X_w, Y_w)/\partial(x_d, y_d)$ is the Jacobian matrix of $(X_w, Y_w)$. Then we have

$$\begin{aligned} \|\Delta\boldsymbol{P}_w\| &\approx \left\| \frac{\partial(X_w, Y_w)}{\partial(x_d, y_d)} \Delta\boldsymbol{p}_d \right\| \\ &\leq \left\| \frac{\partial(X_w, Y_w)}{\partial(x_d, y_d)} \right\|_F \|\Delta\boldsymbol{p}_d\| \end{aligned} \quad (9)$$

where $\|\cdot\|_F$ is the Frobenius norm. Here, homography error is defined as $HE(x_d, y_d) = \left\| \frac{\partial(X_w, Y_w)}{\partial(x_d, y_d)} \right\|_F$. We calculate the homography error by numerical differentiation, specifically compute the average displacement of the eight neighboring pixels of a target pixel $(x,y)$:

$$HE(x,y) \approx \frac{1}{8} \sum_i \|\Delta\boldsymbol{P}_i\|, \quad (10)$$

as illustrated in Figure 5.




$$HE(x,y) \approx \frac{\sum_1^8 \|\Delta P_i\|}{8}$$

Figure 5: Homography error estimation by numerical differentiation. The estimation uses the distance between pixels when projecting adjacent pixels in camera coordinates onto barn coordinates.

## 3.2 Online Steps

**(i) Cow Detection Using Rotated Bounding Boxes.** Rotated bounding boxes can be detected using detectors such as YOLOV5 (Hu, 2020) or YOLOV8 (Ultralytics, 2024). In this method, we track dairy cows using rotated bounding boxes as they offer tight bounding box detection for dairy cows.

**(ii) Selecting Reliable Bounding Boxes Using Homography Error.** The bounding box to be projected onto the barn coordinates is selected using the homography error obtained in 3.1. When the center coordinates of the detection bounding box is $\boldsymbol{p}_c = (x_c, y_c)^T$, the bounding box for which $HE(\boldsymbol{p}_c) < T_{HE}$ is adopted as a bounding box with high reliability. $T_{HE}$ is the threshold of homography error.

**(iii) Inter-Camera Individual Association.** This step integrates the individual dairy cows that cross multiple cameras. Figure 6 shows the flow from the perspective transformation to the integration. Using the homography transformation and the plane projection transformation parameters obtained earlier, bounding box $R$ is projected onto common barn coordinates. In this case, $R = (x, y, w, h, \theta)$ is a bounding box with a center point $(x,y)$, width $w$, height $h$, and rotation angle $\theta$; the four vertices of $R$ are $\boldsymbol{r}_l = (x \pm (w\cos\theta - h\sin\theta)/2, y \pm (w\cos\theta - h\sin\theta)/2)(l = 1,2,3,4)$. The points projected onto barn coordinates are represented by rectangle $\boldsymbol{R}'_l = Mf(\boldsymbol{r}_l; \boldsymbol{k})$.

Let $R'_a$ and $R'_b$ be the rectangles of $R_a$ and $R_b$ projected onto the barn coordinates. If Intersection over Union $\mathrm{IoU}(R'_a, R'_b) = |R'_a \cap R'_b| / |R'_a \cup R'_b|$ exceeds a certain threshold, it is determined that $R'_a$ and $R'_b$ correspond to the same individual and are integrated. Specifically, we use the Hungarian method to find a one-to-one correspondence between the rectangle projected from adjacent cameras. In this case, we integrate them as the same individual when $\mathrm{IoU} > 0.1$.

**(iv) Inter-Camera Tracking.** We track cows using bounding boxes integrated in barn coordinates. In this study, we use the SORT method (Bewley et al., 2016) for tracking on the barn coordinates of dairy

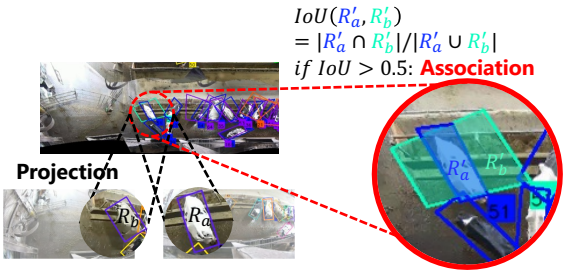



Figure 6: The process from bounding box projection transformation to integration. $R_a$ and $R_b$ indicate the same cow. If the overlap of $R'_a$ and $R'_b$ is higher than 0.5, these bounding boxes are associated with the same cow in barn coordinates.

cows. In order to enable tracking using SORT, we approximate the rectangle $R'$ into bounding box $V$ by taking the average of the inscribed and circumscribed bounding boxes of the rectangle projected onto barn coordinates. By matching this bounding box with the bounding box $V'$ predicted by the Kalman filter, the tracking trajectory in barn coordinates is updated.

# 4 EXPERIMENTS

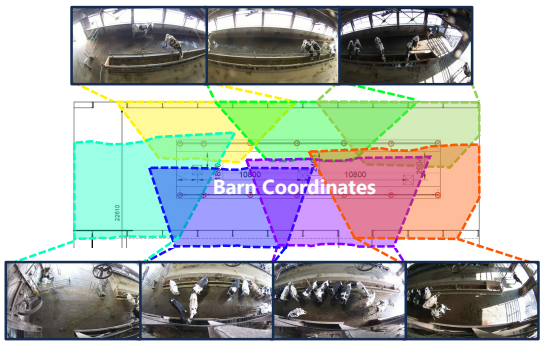
We conducted an experiment to demonstrate the effectiveness of the proposed method. Specifically, we compared the accuracy of the proposed location-based method with that of a feature-based tracking method, with and without homography error (HE).

## 4.1 Datasets

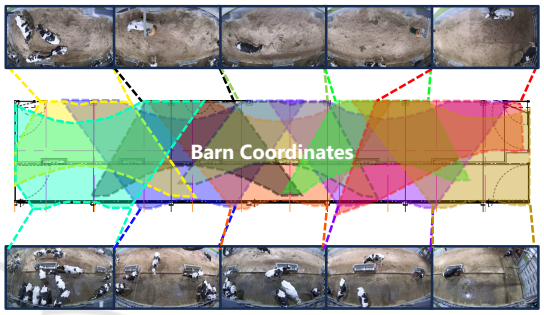
We constructed two video datasets corresponding to major types of barns: free-stall and free-range barns. In free-stall barns, the feeding area and sleeping area are separate, whereas in free-range barn, dairy cows can sleep wherever they like. (Akizawa et al., 2022; Aou et al., 2024) only conducted experiments in the feed area of a free-stall barn, so the effectiveness of the system in environment with dairy cows in various poses, such as a free-range barn, was not confirmed.

The videos were recorded with multi-cameras installed on the ceiling of barns at Obihiro University of Agriculture and Veterinary Medicine. Each dairy cow in the videos was manually annotated with the correct data, including rotated bounding boxes and the individual ID.

**Free-Stall Barn Dataset.** The dataset consists of 47-minute video recorded synchronously by seven barn cameras. The cameras covered the feeding area, as shown in Figure 7a, to comprehensively capture the movement of dairy cows. The adjacent cameras were arranged to partially share their fields of view. The




(a) Free-stall barn




(b) Free-range barn

Figure 7: Camera layout in each barn dataset. The barn coordinates represent the shape of each camera's image when projected onto a panoramic view of the barn.

dataset was split into two subsets: 17-minute videos (15:40 - 15:57) for training the appearance feature extractor, and 30-minute videos (15:57 - 16:27) for evaluating the tracking system. The number of individuals was 63, and the number of bounding box pairs was 20,926.

**Free-Range Barn Dataset.** The dataset consists of 40-minute video recorded synchronously by ten barn cameras. The cameras covered the entire area of a free-range barn, see Figure 7b. The dataset was split into two subsets: 10-minute videos (05:20 - 05:30) for training the feature extractor, 30-minute videos (05:30 - 06:00) for evaluating tracking. The number of individuals was 43, and the number of bounding box pairs was 94,734.

## 4.2 Tracking Metrics

We use Multi-Objects Tracking Accuracy (MOTA) (Bernardin and Stiefelhagen, 2008) and Identification F1 (IDF1) (Ristani et al., 2016) to evaluate tracking accuracy. MOTA focuses on tracking performance, while IDF1 evaluates the degree of agreement at the identification level of the tracking results.

Table 1: Multi-camera individual tracking accuracy of the proposed method compared to the appearance feature-based method. HE indicates the use of homography error for reliable bounding box selection.

| Method | Free-stall barn | | Free-range barn | |
|---|---|---|---|---|
| | MOTA [%] ↑ | IDF1 [%] ↑ | MOTA [%] ↑ | IDF1 [%] ↑ |
| Appearance-Feature Based | 27.5 | 19.7 | 51.5 | 26.5 |
| Appearance-Feature Based + HE | 61.8 | 27.0 | 51.1 | 26.5 |
| (Aou et al., 2024) | 81.2 | 51.0 | 95.2 | 66.9 |
| Ours (Aou et al., 2024) + HE | **86.1** | **66.7** | **95.3** | **69.8** |

Table 2: Comparison of tracking accuracy on videos divided into three segments. Even after splitting the data, the proposed method consistently outperformed the other methods.

(a) Free-stall barn

| Evaluation period (min) | MOTA [%] ↑ | | | | IDF1 [%] ↑ | | | |
|---|---|---|---|---|---|---|---|---|
| | Feature Based | +HE | (Aou et al., 2024) | +HE | Feature Based | +HE | (Aou et al., 2024) | +HE |
| 15:57-16:07 | 33.2 | 67.5 | 81.0 | **84.5** | 33.0 | 46.1 | 62.7 | **76.6** |
| 16:07-16:17 | 25.5 | 58.6 | 80.4 | **87.7** | 31.2 | 39.9 | 69.0 | **81.3** |
| 16:17-16:27 | 25.7 | 58.8 | **86.6** | 84.8 | 36.0 | 41.1 | 76.0 | **81.2** |

(b) Free-range barn

| Evaluation period (min) | MOTA [%] ↑ | | | | IDF1 [%] ↑ | | | |
|---|---|---|---|---|---|---|---|---|
| | Feature Based | +HE | (Aou et al., 2024) | +HE | Feature Based | +HE | (Aou et al., 2024) | +HE |
| 05:30-05:40 | 79.8 | 79.9 | 98.7 | **98.9** | 65.7 | 66.6 | 92.5 | **97.2** |
| 05:40-05:50 | 54.1 | 53.3 | **96.0** | 95.4 | 33.6 | 33.6 | 78.8 | **79.7** |
| 05:50-06:00 | 43.9 | 41.0 | **92.9** | **92.9** | 39.2 | 37.6 | **97.0** | **97.0** |

### 4.3 Experimental Setting

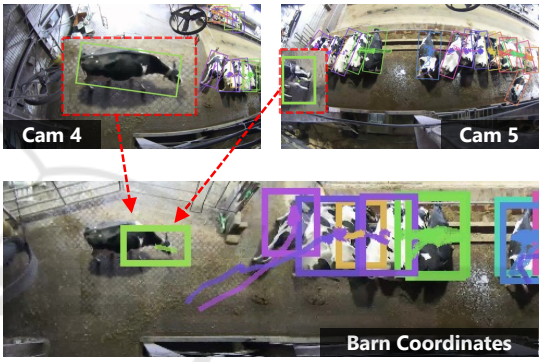
**Appearance Feature Extraction.** We implemented an appearance feature-based tracking system for inter-camera individual association for a comparative experiment. For appearance feature extraction, we used ResNet152 pre-trained on ImageNet (Krizhevsky et al., 2012) and fine-tuned for each dataset using SoftTriple-Loss (Qian et al., 2019).

**Homography Error (HE) Threshold.** For the free-stall barn, $T_{HE} = 0.5$. For the free-range barn, $T_{HE} = 1.0$.

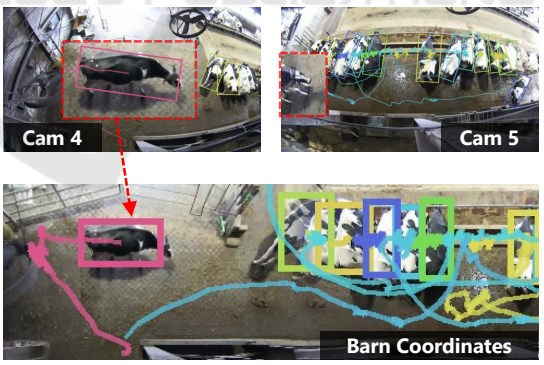
**Camera Setting.** The resolution of the video used was 1,920 × 1,080 pixels, 1fps, and the viewing angle was 124.3°.

## 5 RESULTS AND DISCUSSIONS

**Effectiveness of Homography Error.** To demonstrate the effectiveness of homography error (HE), we compare the accuracy of location-based (Aou et al., 2024) and appearance-based tracking methods with and without HE. As shown in Table 1, the tracking accuracy improved for all methods with bounding box selection and HE. Combining HE with a conventional location-based method(Aou 2024) improved MOTA and IDF1 by 4.9% and 15.7% on the free-stall dataset, respectively. Figure 8 shows an example of successful tracking. As shown in Figure 8, the proposed method with HE succeeded in tracking (Figure 8b), while the conventional method without HE failed due to an ID switch(Figure 8a). This is because, without HE, the same dairy cow in Cam5 are merged, causing greater bounding box misalignment in barn coordinates. In contrast, using HE reduces the bounding box misalignment, as unreliable bounding boxes are excluded.



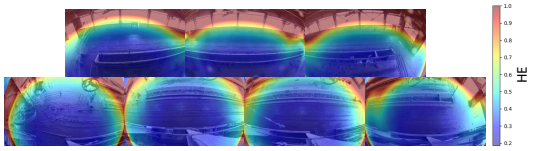

(a) (Aou et al., 2024)



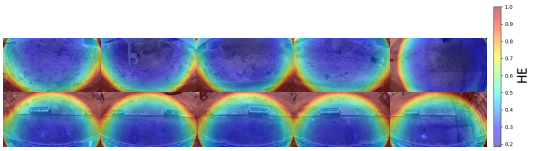

(b) Ours (Aou et al., 2024)+HE

Figure 8: Successful tracking examples when the homography error is used for the free-stall barn dataset. Same color indicates same individual. In our proposed method, the dairy cow indicated by the arrows are continuously tracked.

In addition, as shown in Table 1, HE was effective when combined with an appearance feature based tracking method. MOTA and IDF1 were improved by 34.3% and 7.3% on the free-stall dataset, respectively. When the homography error was not taken into ac-




(a) Free-stall barn




(b) Free-range barn

Figure 9: Visualization of the homography error for each camera, with colors representing the magnitude of the error. In the free-range barn, the homography error shows less variation between cameras compared to the free-stall barn. This is due to smaller differences in the installation angles of the cameras.

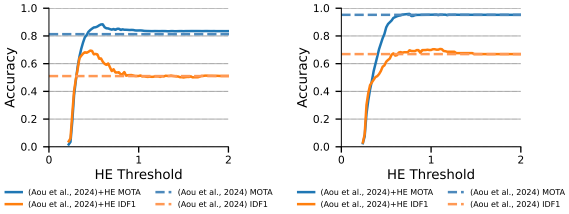
count, there were many cases of integration failure due to the large differences in size and orientation of the dairy cows if the shared fields of view were large, and there were many cases of integration failure. Excluding the bounding boxes with large homography error reduced the difference in feature values for the same cows between cameras, leading to improved tracking accuracy.

To demonstrate the consistency of the proposed method regardless of the input, we report the tracking accuracy for three video segments. We divided the 30-minute tracking evaluation video into three 10-minute segments. As shown in Table 2, the proposed method with HE consistently achieved better results than comparative methods for both datasets.

**Robustness to Camera Installation Angles.** As shown in Table 1, the accuracy improvement was greater for the free-stall dataset than the free-range dataset. This is because, as shown in Figure 9, the camera installation angles in the free-range dataset were aligned, and the HE was small compared to those in the free-stall dataset. These results indicate that the proposed method is robust against changes in camera installation angles.

As shown in Table 2b, the proposed method improved tracking accuracy on the free-range dataset during the first segment (05:30-05:40), when the dairy cows were more active, compared to the second and third segments (05:40-06:00), where their movement was less pronounced. This result suggests that the proposed method reduced the number of ID switches caused when the cows moved across the cameras' fields of view.

**Tracking Accuracy for Varying HE Threhsold.** Figure 10 shows the tracking accuracy with various HE thresholds ($T_{HE}$) for the two datasets. The accuracy was the highest when the threshold was $T_{HE} = 0.5$ for the free-stall dataset and $T_{HE} = 1.0$ for the free-range dataset. Increasing the HE threshold expands the excluded areas indicated by the brown and yellow shading in Figure 9. if the threshold value is set too low in either of the data sets, the tracking accuracy decreases as too many bounding boxes are excluded. One of the limitations of our proposed method is that it requires careful tuning of HE threshold.




(a) Free-stall barn (b) Free-range barn

Figure 10: The relationship between the homography error (HE) threshold and tracking accuracy.

Table 3: Tracking accuracy using homography error (HE) compared with using angle of view (AoV). We compared our proposed method utilizing HE with utilizing AoV.

(a) Free-stall barn

| Method | MOTA | IDF1 |
|---|---|---|
| (Aou et al., 2024) | 81.2 | 51.0 |
| (Aou et al., 2024) + HE | **86.1** | **66.7** |
| (Aou et al., 2024) + Angle of View | 84.8 | 54.7 |

(b) Free-range barn

| Method | MOTA | IDF1 |
|---|---|---|
| (Aou et al., 2024) | 95.2 | 66.9 |
| (Aou et al., 2024) + HE | 95.3 | **69.8** |
| (Aou et al., 2024) + Angle of View | **95.7** | 68.1 |

**Homography Error vs. Angle of View.** To demonstrate the effectiveness of HE estimation, we compared the tracking accuracy when reliable bounding boxes were selected using HE (proposal) and when they were simply narrowed by decreasing the angle of view (AoV).

As shown in Table 3a, MOTA and IDF1 using HE were 1.3% and 12.0% higher than those using AoV on the free-stall dataset. While the AoV does not account for the varying capture angles of the cameras, the homography error evaluates the error for each camera, allowing for the establishment of exclusion areas specific to each camera. As shown in Table 3b, the accuracy improvement with the free-range dataset was relatively small compared to the free-stall dataset; IDF1 using HE was 1.7% higher than AoV, while MOTA was slightly lower. The reason is that for the free-range dataset the shooting angles of cameras were nearly top-down and aligned, as shown in Figure 9, and thus, the effect of HE was relatively small.

## 6 CONCLUSIONS

In this paper, we proposed a method for the multi-camera tracking of dairy cows that utilizes the homography error for selecting reliable bounding boxes. Experiments on actual scenes showed that the proposed method achieved high accuracy for two different barn environments; even when the camera shooting angles were unaligned, the tracking accuracy was improved by appropriately selecting reliable bounding boxes using the homography error. In addition, experiments confirmed that evaluating homography error is effective for appearance-feature based tracking methods, not just location-based methods.

Future work includes verifying our method can handle different camera configurations and arrangement.

## ACKNOWLEDGEMENTS

The authors thank the members of Obihiro University of Agriculture and Veterinary Medicine and Tsuchiya Manufacturing Co. Ltd for helpful discussions and for providing the video data of barn.

## REFERENCES

Akizawa, K., Yamamoto, Y., and Taniguchi, Y. (2022). Dairy cow tracking across multiple cameras with shared field of views. Technical Report 299, The institute of image electronics engineers of Japan. (in Japanese).

Aou, S., Yamamoto, Y., Nakamura, K., and Taniguchi, Y. (2024). Multi-camera tracking of dairy cows using rotated bounding boxes. In *Proc. IWAIT*.

Bernardin, K. and Stiefelhagen, R. (2008). Evaluating multiple object tracking performance: the clear mot metrics. *EURASIP journal on image and video processing*, 2008:1–10.

Bewley, A., Ge, Z., Ott, L., Ramos, F., , and Upcroft, B. (2016). Simple online and realtime tracking. In *Proc. ICIP*.

Bredereck, M., Jiang, X., Körner, M., and Denzler, J. (2012). Data association for multi-object tracking-by-detection in multi-camera networks. In *2012 Sixth International Conference on Distributed Smart Cameras (ICDSC)*, pages 1–6.

Derek, G., Zachár, G., Wang, Y., Ji, J., Nice, M., Bunting, M., and Barbour, W. W. (2024). So you think you can track? In *Proc. WACV*.

Eshel, R. and Moses, Y. (2008). Homography based multiple camera detection and tracking of people in a dense crowd. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8.

Fleuret, F., Berclaz, J., Lengagne, R., and Fua, P. (2008). Multicamera people tracking with a probabilistic occupancy map. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):267–282.

Hu, K. (2020). YOLOV5 for oriented object detection. https://github.com/hukaixuan19970627/yolov5_obb.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.

Mar, C. C., Zin, T. T., Tin, P., Honkawa, K., Kobayashi, I., and Horii, Y. (2023). Cow detection and tracking system utilizing multi-feature tracking algorithm. *Scientific reports*, 13(17423).

Qian, Q., Shang, L., Sun, B., Hu, J., Li, H., and Jin, R. (2019). Softtriple loss: Deep metric learning without triplet sampling. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6450–6458.

Ristani, E., Solera, F., Zou, R., Cucchiara, R., and Tomasi, C. (2016). Performance measures and a data set for multi-target, multi-camera tracking. In *Proc. ECCV*, pages 17–35. Springer.

Ristani, E. and Tomasi, C. (2018). Features for multi-target multi-camera tracking and re-identification. In *Proc. CVPR*, pages 6036–6046.

Shirke, A., Saifuddin, A., Luthra, A., Li, J., Williams, T., Hu, X., Kotnana, A., Kocabalkanli, O., Ahuja, N., Green-Miller, A., Condotta, I., Dilger, R. N., and Caesar, M. (2021). Tracking grow-finish pigs across large pens using multiple cameras. *arXiv preprint arXiv:2111.10971*.

Ultralytics (2024). Oriented Bounding Boxes Object Detection. https://docs.ultralytics.com/ja/tasks/obb/.

Yamamoto, Y., Akizawa, K., Aou, S., and Taniguchi, Y. (2024). Entire-barn dairy cow tracking framework for multi-camera systems. *Computers and Electronics in Agriculture*, 229.

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE transactions on pattern analysis and machine intelligence*, 22(11):1330–1334.