DISTORTION CORRECTION AND STITCHING OF OVERLAPPING CATTLE BARN IMAGES

Risako Ii† Ryosuke Furuta†† Yukinobu Taniguchi†
†Tokyo University of Science ††The University of Tokyo

ABSTRACT

With the increase in the scale of dairy firms and the popularity of stall-free barns, the management of individual dairy cows is becoming more difficult. The obvious solution, installing wide-angle cameras on the ceiling, faces difficulties in grasping the positions of the cows in the barn as the number of cameras increases due to distortion and the overlapping of the captured images. In this study, we aim to create a panoramic image that satisfies the following requirements: the dairy cows must not be cropped, duplicated or missed, and the final image must be effectively seamless. We propose a method that extracts the individual regions of the dairy cows and add them to an underlying panoramic image. We conduct a user evaluation experiment and compare the proposed method with conventional methods such as multi-screen displays and a simple composition method.

Keywords: Dairy cow, Panoramic image, Distortion correction

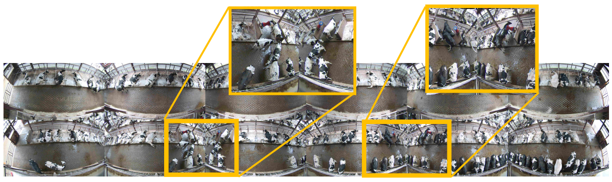


Figure 1: Example of multi-screen display. The source images are distorted and some cows are duplicated in the images.

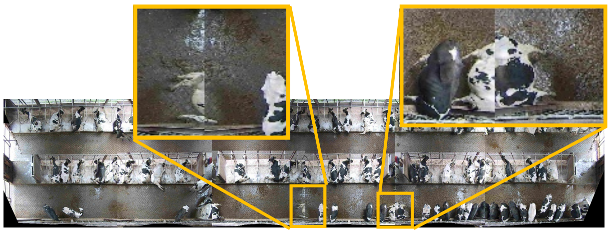


Figure 2: Example of panoramic simple composite image generated from source images.

1. INTRODUCTION

With the continued decrease in dairy farmer numbers, the scale of each dairy firm continues to increase. Moreover, there has been a shift from the tethered system to the stall-free system, which requires less labor and reduces stress on the dairy cows. However, the stall-free approach has a problem in that it makes it difficult for farmers to observe each cow 24 hours a day. Although installing wide-angle cameras on the ceiling yields cost-effective monitoring, it is still difficult to grasp cow position as the increase in barn size increases the number of cameras needed.

To solve this problem, we need to generate a bird's eye view of the entire barn. Figure 1 shows a multi-screen display that shows images from cameras installed on the ceiling of the barn (hereinafter referred to as source images). As shown in Figure 1, the source images are distorted and dairy cows are duplicated because a single cow can be captured by different cameras at the same time.

To improve the efficiency of overall barn management this study proposes a method for generating a panoramic composite image from multiple source images. Specifically, we aim to create panoramic images that satisfy the following three requirements: (a) no dairy cow can be cropped, (b) each cow must be shown without duplication or omission, (c) seams in the composite images must not be noticeable. To satisfy these three requirements, the proposed method extracts the individual regions of dairy cows, and adds them to a panoramic image of the barn empty.

2. RELATED WORK

The existing approach to fusing multiple images is as follows : (1) estimate the extrinsic camera parameters from neighboring images, (2) apply projective transformations to align all the images on a compositing plane, (3) deciding which pixels to use in the areas of image overlap. Step (3) is required to avoid ghosts caused by moving objects in overlap areas. The conventional method for de-ghosting is to simply select the pixel closest to the center of each image from among the overlapping pixels. However, as shown in Figure 2, there are the problems in that the seams are very noticeable and some cows are cropped, which means requirements (a) and (c) cannot be satisfied.

Szeliski et al. [1] proposed a method to build a graph with the regions of difference (ROD) between neighboring images as vertices, and remove all but one difference region. Charles et al. [2] proposed a method for pixel selection that uses feature point matching to matches detected objects between neighboring images. In our case, the distance between neighboring cameras is so large that a cow in an overlap region significantly changes in appearance between cameras, for example, different sides of the cow body appear. Therefore, it is difficult to extract ROD and to match the same dairy cows, which may not satisfy requirements (a) and (b).

Different from the methods of Szeliski et al. [1] and Charles et al. [2], the proposed method considers object depth, that is cow height, because the distance between the sources and the cows is rather small; previous methods [1][2] assume that the objects are far from the camera. A multi-view stereo method such as SfM-MVS [3] can be applied to generate a panoramic image by projecting a 3D model onto a surface parallel to the floor. However, it does not work because the feature point matching fails due to the change in object appearance between neighboring cameras.

3. PROPOSED METHOD

In this study, we propose to use a cattle barn floor map (hereinafter referred to as the barn map) as the compositing surface and use the results of individual cow region extraction to create a panoramic image, without the complication of 3D reconstruction by assuming an average cow height. As shown in Figure 3, the proposed method consists of three steps: initial settings, generating an empty barn image, and compositing barn panoramic images.

3.1. Initial setting

We perform camera calibration to obtain the camera parameters and lens distortion coefficients. Assuming a pinhole camera model with lens distortion,

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

where $\boldsymbol{P} = (X, Y, Z, 1)^{\mathrm{T}}$ is the homogeneous coordinate of a 3D point in the barn, $\boldsymbol{p_d} = (u, v, 1)^{\mathrm{T}}$ is the homogeneous coordinate of a 2D point on the source image, $\boldsymbol{p} = (x, y, 1)^{\mathrm{T}}$ is the homogeneous coordinate after distortion correction, f is the function to correct the distortion, $\boldsymbol{K}$ represents the intrinsic camera parameters, $(\boldsymbol{R}|\boldsymbol{t})$ represents the extrinsic camera parameters, and $\boldsymbol{k}$ represents the distortion coefficients. Parameters $\boldsymbol{K}$, $(\boldsymbol{R}|\boldsymbol{t})$, $\boldsymbol{k}$ are computed from multiple pairs of corresponding 3D point $\boldsymbol{P}_i = (X_i, Y_i, Z_i)^{\mathrm{T}}$ in the barn space and 2D point $\boldsymbol{p}_i = (x_i, y_i)^{\mathrm{T}}$ in the barn image; the point correspondences are manually specified. The projective transformation from the source image to the barn map, i.e., the floor plane $Z = 0$ in the barn space, can be written as $(X, Y, 1)^{\mathrm{T}} \sim \boldsymbol{M}(x, y, 1)^{\mathrm{T}}$, where $\boldsymbol{R} = (\boldsymbol{r_1}, \boldsymbol{r_2}, \boldsymbol{r_3})$ and $\boldsymbol{M} = (\boldsymbol{K}(\boldsymbol{r_1}, \boldsymbol{r_2}, \boldsymbol{t}))^{-1}$.

3.2. Generating empty barn panoramic image

We first generate a panoramic image of the barn empty as the underlying image for cow region projection. First, we capture the source images when no cow is present, apply distortion correction and projective transformation using the camera parameters obtained in the initial setting step, and then composite the images to generate an empty barn panoramic image. The entire barn is composed by applying the method of simple composition to overlapping regions.

3.3. Compositing barn panoramic images

We extract the regions of individual dairy cows with Mask R-CNN[4], and apply distortion correction and projective transformation to each extracted region. Note that, on the projected image, cows appear larger than the proper size on the floor as shown in Figure 4. That is because the back of a standing cow is higher than the floor, while the compositing surface is the floor. Therefore, we apply projective transformation $(X, Y, 1)^{\mathrm{T}} \sim \boldsymbol{M'}(x, y, 1)^{\mathrm{T}}$ to the regions of standing cows, where $\boldsymbol{M'} = (\boldsymbol{K}(\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{t} + h\boldsymbol{r_3}))^{-1}$ and h is average cow height. We call this transformation scale correction, because applying projective transformation with matrix $\boldsymbol{M'}$ is equivalent to applying projective transformation with matrix $\boldsymbol{M}$ and resizing by a factor of μ, where $\mu = 1 - \frac{h}{i_3}$ and $(i_1, i_2, i_3)^{\mathrm{T}} = -\boldsymbol{R}^{\mathrm{T}}\boldsymbol{t}$.

We generate a panoramic image by compositing the cow regions onto the empty barn image. To avoid duplicates, only the cow region closest to the image center is used in compositing.

The process of matching cow regions is as follows: Let M_a and M_b be cow regions extracted from neighboring source images a and b, respectively. We decide M_a and M_b are the same cow if $\mathrm{IoU}(M_a, M_b)$ exceeds a predetermined threshold, where

$$\mathrm{IoU}(M_a, M_b) = \frac{Area(M_a \cap M_b)}{Area(M_a \cup M_b)},$$

which represents the overlap between M_a and M_b. The threshold is needed because, when the cows are close to each other, the cow regions may overlap even if they are different cows.

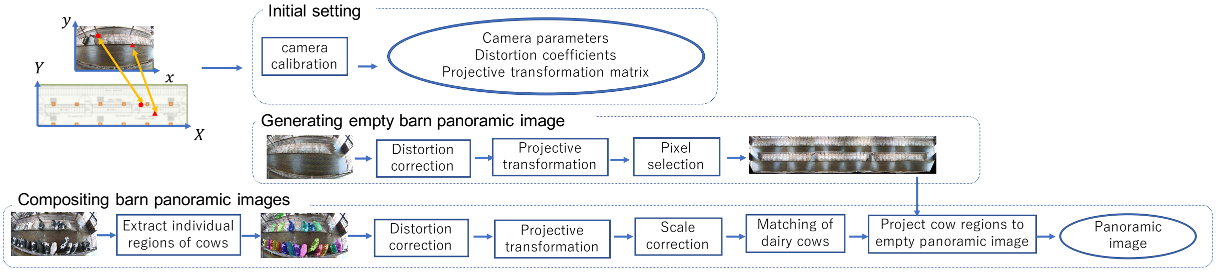


Figure 3: Flowchart of the proposed method

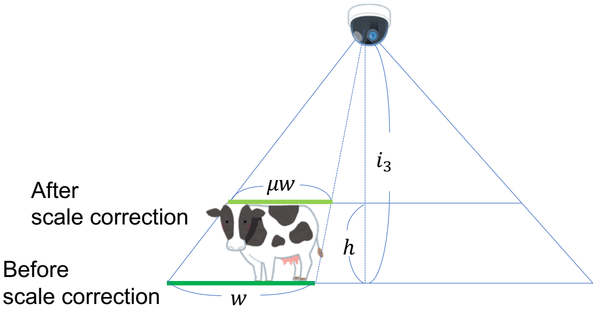


Figure 4: Differences in extent of the cow region

Figure 5: Panoramic image generated by the proposed method (without scale correction)

Figure 6: Panoramic image generated by the proposed method (with scale correction)

4. EXPERIMENTS

We evaluate the accuracy of cow matching in Sec. 4.2. In Sec. 4.3, we examine the subjective quality of the composite panoramic image by user evaluation in order to confirm the effectiveness of the proposed method.

4.1. Experimental conditions

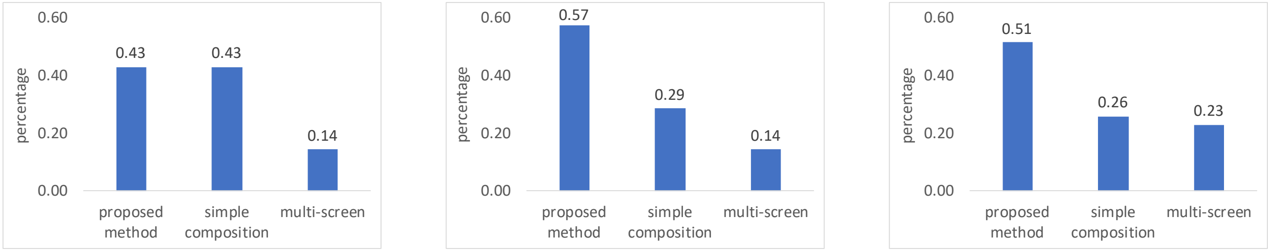
In the experiments, we used images taken from 12 cameras installed on the ceiling of a barn. The viewing range of the cameras was about 6.4 $[m] \times$ 21.6 $[m]$. The cameras were installed at 10.8 [m] intervals with an overlap of about 3.6 $[m]$ between adjacent cameras. In terms of camera calibration, 24 corresponding points on average were manually assigned to each image. Dairy cow matching was applied to just the dairy cow regions detected with Mask R-CNN. Average cow height was set to 1.5 $[m]$. The IoU threshold was set to 0.1.

4.2. Experiment1: Evaluation of accuracy of dairy cow matching

We generated 30 panoramic images at consecutive times to assess the accuracy of cow matching. Figures 5 and 6 show examples of the panoramic images generated by the proposed method without and with scale correction. The scale correction was applied to only cow regions in the aisle area, in which most cows are standing. As evaluation metrics, we used average precision (AP) and the average recall (AR). For each of the 300 pairs of overlapping areas in neighboring source images, we evaluated the precision and the recall, and their averages were determined. Scale correction was applied to only standing cows.

Table 1 shows the results with and without scale correction. This shows the scale correction that takes account of the depth is effective in that AP and the AR were improved.

As analysis of failure cases found there were two main causes of failure: the first is the error in the projective trans-



(a) ease of following (b) ease of finding the current position (c) ease of seeing the panoramic image

Figure 7: Results of user evaluation

Table 1: Average precision (AP) and recall (AR) of cow matching compared to the results with and without scale correction

	AP	AR
with scale correction	0.628	0.339
without scale correction	0.400	0.083

formation, which may cause the matching failure due to misalignment of the cows; the second is cows whose regions overlapped only at the feet, which may cause matching failure because of small IoU.

4.3. Experiment2: User evaluation experiment of panoramic video

We created 10 second videos by arranging the panoramic images composited by the following three methods : (1) the proposed method (Figure 6), (2) simple composition (Figure 2), and (3) multi-screen display (Figure 1). We conducted user evaluations, Web surveys, that involved 35 male and female participants raging in age from 20 to 50. After watching the videos, we asked the participants to rank them in terms of the following three items : (i) ease of following each cow, (ii) ease of finding the current position of the cow, and (iii) ease of seeing the panoramic image.

Figure 7 shows the percentage of participants who selected the video with the highest ratings for each item. Figure 7(a) shows that the proposed method yielded the same ease of following each cow as simple composition. This was probably due to matching failure which caused duplicated, missed, and overlapped cow regions making it difficult to follow individual cows. Figures 7(b) and (c) show that the proposed method is the best in the other two evaluation metrics. As there are no overlapping areas and the cows are projected one by one in the proposed method, the seams were evaluated to be less noticeable than those created by the compared methods, so requirement (c) is considered to be satisfied.

5. CONCLUSION

In this study, we proposed a method for generating panoramic images of complete barns to facilitate the management of dairy cows. The results of a user evaluation show that the proposed method yields more intuitive results than the conventional methods. The seams in the composite image were less noticeable. Indicative of the limitation of the proposed method, requirements (a) and (b) were not fully achieved. In particular, there were many duplicated and missing cows, so improving the accuracy of cow matching will be the goal of future research. Possible solutions are to improve the accuracy of the camera calibration to minimize the error in the projective transformation, and to take into account the distance between cows as well as the area of overlapping region when matching. In addition, because our method assumes that all cows on the aisle are standing, we need to reconsider how to decide which cow is standing.

ACKNOWLEDGMENT

We would like to thank Tsuchiya manufacturing Co. Ltd. and Obihiro University of Agriculture and Veterinary Medicine for providing the images of the cattle barn. This work was supported by the Strategic Core Technology Advancement Program of METI.

REFERENCES

[1] Matthew Uyttendaele, Ashley Eden, and Richard Szeliski, Eliminating ghosting and exposure artifacts in image mosaics, In CVPR, Vol. 2, pp. II–II, (2001).

[2] Charles Herrmann, Chen Wang, Richard Strong Bowen, Emil Keyder, and Ramin Zabih, Object-centered imagestitching, In ECCV, pp. 821-835, (2018).

[3] Steven M.Seitz and Brian Curless, James Diebel, and Daniel Scharstein, and Richard Szeliski, A comparison and evaluation of multi-view stereo reconstruction algorithms, In CVPR, Vol. 1, pp. 519-528, (2006).

[4] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross-Girshick, Mask R-CNN, In ICCV, pp. 2961-2969, 2017.