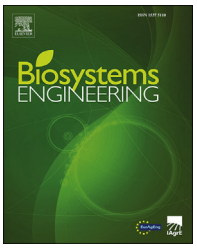
Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/issn/15375110



Special Issue: Biosystems and Metrology
Research Paper

2-D/3-D fusion-based robust pose normalisation of 3-D livestock from multiple RGB-D cameras

Jie Lu [a], *Hao Guo* [a,*], *Ao Du* [a], *Yang Su* [a], *Alexey Ruchay* [c,d], *Francesco Marinello* [b], *Andrea Pezzuolo* [b]

[a] *College of Land Science and Technology, China Agricultural University, Beijing 100083, China*
[b] *Department of Land, Environment, Agriculture and Forestry, University of Padova, Viale dell'Universitá 16, 35020 Legnaro PD, Italy*
[c] *Federal Research Centre of Biological Systems and Agro-technologies of the Russian Academy of Sciences, Orenburg 460000, Russia*
[d] *Chelyabinsk State University, Chelyabinsk 454001, Russia*

ARTICLE INFO

Article history:
Published online xxx

Keywords:
Precision livestock farming
Body measurement
2-D/3-D fusion
Pose normalisation
point clouds

Recently, 3-D scanning data are becoming increasingly important for precision livestock farming. In particular, RGB-D (red, green, blue – depth) data of livestock have come to play a critical role in the field of livestock body measurement. However, the latest livestock pose normalisation methods rely on purely 3-D geometric data and are therefore prone to errors due to noise and missing data. To achieve adequate performance, particularly for different livestock species in practical applications a 2-D/3-D fusion-based robust livestock pose normalisation method is proposed. Firstly, based on advanced 2-D object detection techniques, the proposed approach makes the best use of 2-D information to determine the accurate orientation of livestock in 3-D. Secondly, the 2-D detection results are used to generate frustums in 3-D space to locate livestock targets, which markedly reduces the search space and improves segmentation. Finally, based on a bilateral symmetry-based pose normalisation framework, a more robust pose normalisation algorithm is applied. Compared to existing pose normalization methods that operate in 3-D, extensive experiments with multiple view RGB-D data of livestock show that the proposed method is more robust and practical than existing methods. The proposed algorithm provides pose normalisation in an automatic body measurement system for livestock. This study proposes that the idea that 2-D/3-D fusion-based strategies in 3-D should be explored in more detail, particularly for cases in which the 3-D input captured by consumer designed RGB-D cameras which are often noisy and miss values at certain pixels. All the training databases and codes used in the study can be downloaded freely.

* *Corresponding author.*
E-mail address: guohaolys@cau.edu.cn (H. Guo).
https://doi.org/10.1016/j.biosystemseng.2021.12.013
1537-5110/

Nomenclature

2-D/2D	Two dimension
3-D/3D	Three dimension
RGB-D	Red, green, blue and depth
R−CNN	Regions with convolutional neural networks features
CCS	Canonical coordinate system
PN	Pose normalisation
YOLO	You only look once
YOLO v4	Optimised YOLO
COCO	Common objects in context
SPNA	Pose normalisation approach Guo et al. (2019)
OSPNA	Optimised SPNA
mAP	Mean average precision
IoU	Intersection over union
BCS	Body condition score
FDD	Forward direction determination
LS	Livestock segmentation
SLS	Supported livestock species

1. Introduction

The shape of an animal is a significant indicator of its physical condition (Klingenberg, 2010; Savin et al., 2011). In particular, in the field of precision livestock farming, livestock body measurements have become the most direct method to quantify the shape of individual animals (Guo et al., 2019). This study is mainly based on a previous study (Guo et al., 2019) but here the deficiencies of pose normalisation, such as livestock segmentation and forward direction estimation, have been reduced greatly improving the results of pose normalisation.

Several hardware and software technologies have been developed to support measurements and estimation of parameters. Machine vision has been widely used for contactless animal body measurements to overcome the limitations of traditional manual measurement methods (Kuzuhara et al., 2015). Researchers have demonstrated the feasibility of determining the body shape of cattle, pigs, sheep, and horses via videos and images (Azzaro et al., 2011; Brandl & Jørgensen, 1996; Marchant, Schofield, & White, 1999; Menesatti et al., 2014; Pallottino et al., 2015; Tasdemir, Urkmez, & Inal, 2011; Wongsriworaphon, Arnonkijpanich, & Pathumnakul, 2015; Yilmaz, Cemal, & Karaca, 2013). However, there are many challenges in using only the 2-D information contained in images or videos to determine the shape of livestock. These are mainly related to the difficulty of estimating the volume of the animal (Mortensen, Lisouski, & Ahrendt, 2016; Salau, Haas, Junge, & Thaller, 2017a).

Researchers have therefore begun to focus on the use of three-dimensional (3D) information (Li et al., 2020; Zhang, Shen, Wang, Kong, & Yin, 2018). Over the past few decades, many studies have been published contributing to 3-D point cloud processing (Vosselman, Coenen, & Rottensteiner, 2017; Wang, Hu, Wu, & Wang, 2016), autonomous vehicle navigation (Yang et al., 2017), forest assessment (Polewski, Yao, Heurich, Krzystek, & Stilla, 2017), building information modelling (BIM) (Thomson & Boehm, 2015) and industrial applications (Luhmann, 2010). Recently, the consumer depth sensors, such as Intel RealSense or Microsoft Kinect, have provided 3-D data and provided new methods to capture the shape of individual livestock (Hao et al., 2017; Kawasue, Ikeda, Tokunaga, & Harada, 2013; Pezzuolo, Guarino, Sartori, González, & Marinello, 2018; Salau, Haas, Junge, & Thaller, 2017b; Viazzi et al., 2014). Song, Bokkers, van der Tol, Koerkamp, and van Mourik (2018) reported on animal weight estimation errors and the influence of different sources on results. Shuai et al. (2020) presented a 3-D surface reconstruction and body measurement system based on multiple RGB-D cameras. A Kinect depth camera was used to obtain the point cloud of the freely walking pig from three different views and to accurately measure relevant parameters of the livestock.

Thus, with the improvement of optical technologies (Lichti, Qi, & Ahmed, 2012), it has become easier to obtain 3-D information about objects. Many problems that are difficult to solve using 2-D information can be solved with 3-D information to describe objects. However, collection of 3-D information is typically slower than 2-D and as such often leads to problems that might be solved or reduced taking advantage of more rapidly obtained 2-D information. By way of example, in Guo et al. (2019) only 3D geometric information was used to determine the forward direction of livestock: in many practical cases and this led to errors in pose normalisation. In Shuai et al. (2020), although 3-D information of body shape was effectively achieved errors as high as 3−5% remained in the estimation of body dimensions. Research studies are required to show how 2-D and 3-D information might be successfully integrated in order to converge towards more repeatable and less uncertain results.

There are many studies that use 3-D information to detect and segment objects, and use this to achieve a specific goal (Cao et al., 2020; Gong et al., 2020). However, when segmenting a 3-D object, researchers more commonly use 2-D object detection technology to detect objects in 2-D images, thereby reducing the 3-D search space and computation required for 3-D data processing (Qi, Liu, Wu, Su, & Guibas, 2018; Shen et al., 2020). Lahoud and Ghanem (2017) used the latest 2-D object detection technology to make full use of 2-D information to reduce the 3-D search space. Indeed, with the development of artificial intelligence, deep learning technology has rapidly improved, and it is being applied to enhance a wide range of applications in 2-D image detection. 2-D image detection technology combined with deep learning approaches, such as R−CNN (Regions with convolutional neural networks features) (Girshick, Donahue, Darrell, & Malik, 2014), Fast R−CNN (Girshick, 2015), Faster R−CNN (Ren, He, Girshick, & Sun, 2017), and YOLO (Redmon, Divvala, Girshick, & Farhadi, 2016), might then constitute a robust reference for position detection of 2-D image objects (Tassinari et al., 2021), having marked advantages over 3-D detection methods both in terms of time and accuracy. Thus, generally speaking, combining 2-D and 3-D methods can allow more complex practical problems to be addressed (Acharya, Khoshelham, & Winter, 2019; Luo et al., 2020; Xiang, Wang, Lao, & Zhang, 2020).

Please cite this article as: Lu, J et al., 2-D/3-D fusion-based robust pose normalisation of 3-D livestock from multiple RGB-D cameras, *Biosystems Engineering*, https://doi.org/10.1016/j.biosystemseng.2021.12.013

Using multi-view RGB-D data, a 2D/3D fusion-based robust livestock pose normalisation approach was proposed that can take full advantage of 2-D and 3-D information. It was hoped that this can achieve excellent performance, particularly for different livestock species in practical applications. The pose normalisation method proposed to solve the problems encountered during the automatic body measurement of livestock. This method was divided into four components: estimation of the forward direction of the livestock, livestock segmentation, bilateral symmetry plane estimation, and pose normalisation transformation. During forward direction estimation, the latest livestock normalisation method relied on the assumption that the data of livestock must be complete (Guo et al., 2019). However, missing the front or rear of livestock often occurred during 3-D scanning in farming conditions. Thus, using this feature to judge the orientation of livestock will frequently cause errors in forward direction estimation. Similarly, during the step of livestock segmentation, the study of Guo et al. (2019) assumed that an animal stands on flat ground without other farm facilities in the 3-D scene. When there is more than one plane in the point cloud or the ground plane is small, incorrect judgements of the target cluster are common. Finally, pose normalisation transformation is based on the results of the forward direction estimation, target segmentation, and bilateral symmetry plane estimation. Therefore, if the result of the previous processes is incorrect, then the subsequent pose normalisation will fail. The proposed approach solves the deficiencies of forward direction estimation and livestock segmentation. By introducing 2-D object detection technology, a more robust method for the pose normalisation of livestock is proposed. The primary contributions of this paper are as follows:

- An optimised automatic pose normalisation framework is proposed.
- Combination of 2D object detection and 3D point cloud technology to enhance pose normalisation.
- Improved livestock segmentation and estimation of forward direction is demonstrated.
- All training data and codes can be downloaded freely from the provided links (Guo & Lu, 2021) for use by the livestock industry and research community.

2. Materials and methods

2.1. 3-D input requirements

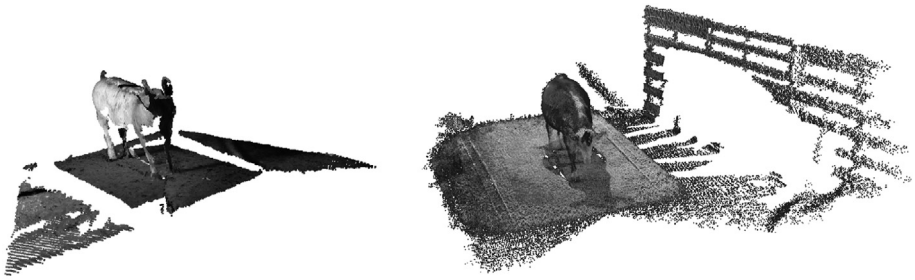
Before starting the research, certain assumptions need to be made and they can be crucial to its success. Firstly, animals are assumed to stand upright on the ground and cannot intersect with other objects. The animal is assumed to stand on the ground with its head facing forward. This means that the top view of animals is almost straight. In addition, the 3-D animal data is point cloud data after registration from different aspects, and thus the integrity of the data can be guaranteed without too much missing. It is important that the data of each animal includes the body, head, and hip. The conditions for obtaining data for each animal should be similar. Point clouds denoted by $S = \{p_i\}$ were used which contain the animal, flat ground, and any obscuring objects in the visual frame. Additionally, to test the robustness and generality of the proposed pose normalisation method, this paper considers both double-view RGB-D data for pigs (Guo et al., 2019) and the triple-view RGB-D data for cattle (Ruchay, Dorofeev, Kalschikov, Kolpakov, & Dzhulamanov, 2019).

The triple-view RGB-D data of cattle were captured from an automated computer vision system. The cattle data capture system is shown in Fig. 1(a). An RGB-D camera was placed on each side of a cattle passage, and the distance between the camera and target was about 2.0 m. An additional camera was placed directly above the passage about 3.0 m above the ground. For more details on the cattle data, please refer to Ruchay, Kober, Dorofeev, Kolpakov, and Miroshnikov (2020). As shown in Fig. 1(b), the double-view RGB-D data of pigs were obtained from two calibrated Xtion Pro RGB-D cameras (ASUS, China) located at the lateral side of pigs. Both multiple RGB-D systems acquired input data that complied with the assumptions of the study. More details of assumptions about the input of 3-D data are given in Guo et al. (2017). Fig. 2 shows examples of the 3-D point cloud data for pigs and cattle captured by these systems. To guarantee that the acquired 3D data complied with the assumptions, only one animal was examined in each image.

(a) Cattle data capture system (Ruchay et al., 2019) (b) Pigs data capture system (Wang et al., 2018)

Fig. 1 – Examples of livestock 3-D scanning system in use.

Please cite this article as: Lu, J et al., 2-D/3-D fusion-based robust pose normalisation of 3-D livestock from multiple RGB-D cameras, *Biosystems Engineering*, https://doi.org/10.1016/j.biosystemseng.2021.12.013



(a) 3-D point cloud data of pig (b) 3-D point cloud data of cattle

Fig. 2 – Example of qualified input multiple view RGB-D data (3-D point clouds).

2.2. 2-D/3-D fusion based pose normalisation

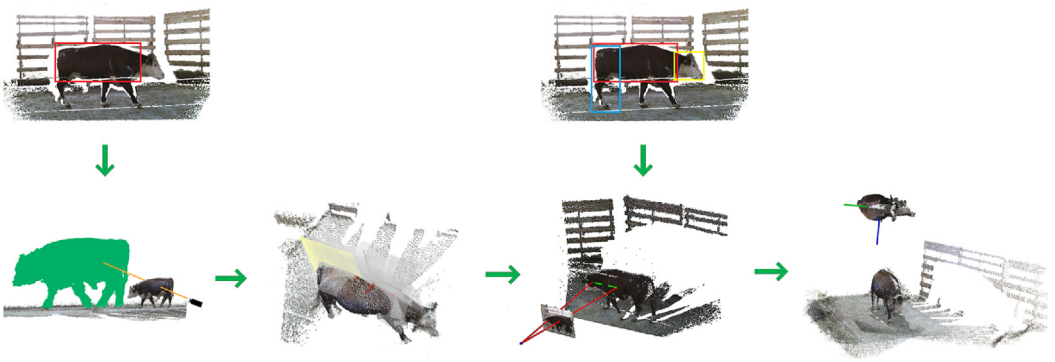
To efficiently extract the body measurements of livestock from 3-D data, it is necessary to separate the animal from the scene. The segmented image of the animal then needs to be place in a predefined canonical coordinate system (CCS) (Sfikas, Theoharis, & Pratikakis, 2011). Therefore, an automatic 3-D point cloud pose normalisation framework is required. The framework proposed can be used for automatic processing of 3-D data of livestock (Guo et al., 2019). However, in practical applications, there are still limitations in performing pose normalisation. In particular, using the algorithm from Guo et al. (2019) missing parts of the animal heads can easily lead to normalisation failure due to an incorrect determination of forward direction. These weaknesses were overcome in this work using a 2-D/3-D fusion strategy, and sections 2.2.2 to 2.2.5 provide more detail. In Fig. 3, a pipeline diagram shows a series of procedures for the proposed method.

The definition of a canonical coordinate system used was that proposed by Guo et al. (2019), and here the canonical coordinate system is only briefly summarised. Fig. 4 shows the details of the CCS for livestock. Multiple RGB-D data of livestock are typically captured in arbitrary orientations and positions in 3-D space. However, the traits of livestock are measured along in one fixed direction. Therefore, it is critical to define a suitable canonical coordinate system for the automatic analysis of 3-D livestock.

2.2.1. Target region detection

The fundamental procedure in the proposed optimisation to obtain the position of the targets, including the head, hip and body, in 2-D. In this study, the existing detection model YOLO v4 (Bochkovskiy, Wang, Liao, & Hao, 2020) was used to detect multiple regions of different sizes in a single 2-D image. The YOLO v4 pretrained model which is trained on the COCO dataset (Lin et al., 2014) was not used because the pretrained model did not contain the targets of interest. Random initialised parameters were used to train the detector to identify the regions of the livestock's head, hip and body in the scene.

To train a detector of YOLOv4, a custom dataset was prepared that contained three types of images: virtual images generated from single-view RGB-D data (pigs and cattle), virtual images generated from double-view RGB-D data (pigs), triple-view RGB-D data (cattle), and RGB images (pigs and



(a) Livestock segmentation. (b) Bilateral symmetry plane estimation. (c) Forward direction estimation. (d) Pose normalization based on rigid transformation.

Fig. 3 – Processing pipeline for proposed pose normalisation method.

Please cite this article as: Lu, J et al., 2-D/3-D fusion-based robust pose normalisation of 3-D livestock from multiple RGB-D cameras, *Biosystems Engineering*, https://doi.org/10.1016/j.biosystemseng.2021.12.013

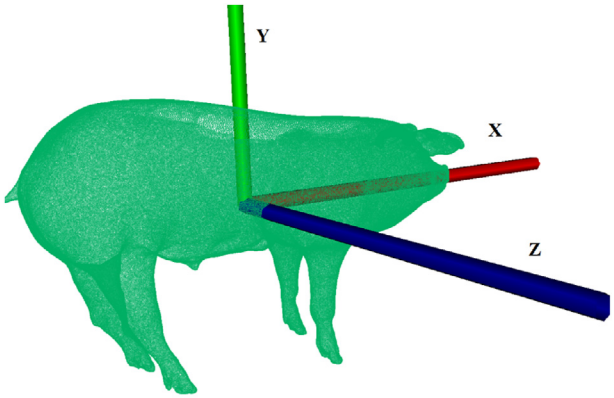


Fig. 4 – Canonical coordinate system definition (Guo et al., 2019).

cattle). Fig. 5 shows an example of the three types of images. Virtual images can be generated from an arbitrary viewpoint in which the virtual image contains as many targets as possible. Additionally, the parameters of the viewpoint were saved to be used to project the 2-D pixel point into 3-D space.

To obtain the forward direction of the livestock, some regions were selected as detection objects. The choice of the region was based on distinct characteristics so that it could be easily obtained through deep learning technology. Therefore, the point of view needed to include the region selected. After comprehensive judgment, three obvious targets were selected: body, hip and head. The three regions selected therefore should contain as much as possible of the body, hip, and head of the livestock. Therefore, there were some uncertain overlaps in these three regions, but this did not affect forward direction estimation. As shown in Fig. 6, the target regions were identified with three colours, and the detected target region is represented by a 2-D window.

Due to the high resolution of RGB information in the image, the target region detection result was more reliable than the 3-D geometry method. The results of the 2-D target region detection were then used to locate the targets in 3-D space. To make the process more efficient, the centre point of the 2-D window was used to represent the position of target in 2-D image. These centre points were then used for livestock segmentation and forward direction estimation.

2.2.2. *Livestock segmentation*

For the Y-axis defined in CCS, it can be inferred that the Y-axis of the livestock will be obtained when the ground plane is estimated. Based on this inference, the same process as in Guo et al. (2019) was used to determine the ground point clouds. In this study, the ground plane point cloud was successfully extracted to obtain the parameters of the ground plane, which supplied the Y-axis of the livestock. Next, planes from the down-sampled data D were extracted and removed and Euclidean cluster extraction was used to divide D into different clusters. The next step was to determine which cluster described livestock. Guo et al. (2019) assumed that the cluster with the largest number of point clouds represented the livestock; however, this assumption is prone to errors in practice when there are large planes or facilities within the scene. In this paper, 2-D object detection technology is introduced to eliminate this assumption and achieve a more robust segmentation.

From section 2.2.1, the position of the target represented by the centre point of the 2-D window was determined. The input image could be an RGB image captured directly or a generated virtual RGB image, and the generated virtual image is typically missing certain targets. Therefore the number of targets that can be detected in the different types of images is also different. With poor-quality images, the required target often cannot be detected. To avoid these problems, to help identify the cluster that describes the livestock, the body of the livestock was chosen as the target because it is easily detected and is rarely missing from images. The 2-D centre point of the target region into 3-D space was then projected. It was assumed from section 2.1, that animals stand upright on the ground and cannot intersect with other objects. Therefore, the cluster intersecting with the projecting ray should be the cluster describing the livestock.

Based on the condition of camera calibration, the camera internal parameter matrix T_i and external parameter matrix T_e required determination when the image was generated to project the points from 2-D space into 3-D space. From section 2.2.1, these parameters were saved when the image was

(a) Virtual images generated from the double-view RGB-D data

(b) RGB images

(c) Virtual images generated from the single-view RGB-D data

Fig. 5 – Types of the 2-D images (pig reference).

Please cite this article as: Lu, J et al., 2-D/3-D fusion-based robust pose normalisation of 3-D livestock from multiple RGB-D cameras, *Biosystems Engineering*, https://doi.org/10.1016/j.biosystemseng.2021.12.013

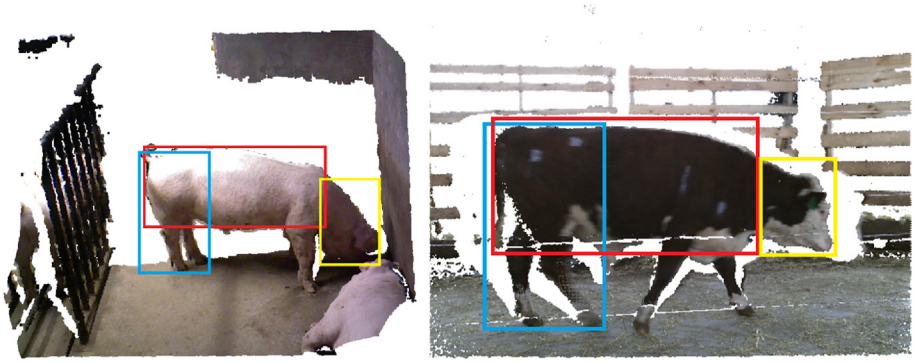


Fig. 6 – Regions of the livestock's body, hip, and head are labelled in red, blue, and yellow. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

generated. Then, using following equation (Andrew, 2001), 3-D points can be projected into the 2-D pixel points:

$$Z_c\begin{bmatrix}u\\ v\\ 1\end{bmatrix}=\begin{bmatrix}f_x & 0 & u_0 & 0\\ 0 & f_y & v_0 & 0\\ 0 & 0 & 1 & 0\end{bmatrix}\begin{bmatrix}R & T\\ 0 & 1\end{bmatrix}\begin{bmatrix}X\\ Y\\ Z\\ 1\end{bmatrix}, \quad (1)$$

$$T_i=\begin{bmatrix}f_x & 0 & u_0 & 0\\ 0 & f_y & v_0 & 0\\ 0 & 0 & 1 & 0\end{bmatrix}, T_e=\begin{bmatrix}R & T\\ 0 & 1\end{bmatrix}$$

where Z_c represents the distance from the camera to the target; u, v are the pixel coordinates in the pixel coordinate system; X, Y and Z are the 3-D coordinates in the world coordinate system; f_x, f_y are the focal lengths in the corresponding directions respectively; u_0 and v_0 are the coordinates of the optical center in the pixel coordinate system; R is the rotation matrix; and T is the translation matrix.

All of these parameters were obtained when the image was generated. The target regions were obtained through 2-D object detection and thus the centre point (u, v) of the target region was obtained in 2-D. Using Eq. (1), the 2-D pixel points (u, v) into 3-D points (X, Y, Z) were projected by reversing the formula. In this study, camera internal parameters T_i were primarily used to convert the camera coordinate system into the pixel coordinate system, and camera external parameters T_e were used to convert the world coordinate system into the camera coordinate system. Fig. 7 shows an example of the projection of 2-D pixels into 3-D space. Thus from Eq. (1), the position of the target in 3-D space was obtained.

The location of the animal's body, which is denoted by $\mathbf{P}_{body}$, was used to determine the cluster that describes the animal. For each cluster C_*, if $\{\mathbf{P}_{body} \in C_*\}$, the cluster that describes the livestock was determined. Fig. 8 illustrates the judgement process for livestock segmentation.

2.2.3. *Bilateral symmetry plane estimation*

Bilateral symmetry plane estimation method was described by Guo et al.(2019), and it is only briefly summarise in this study. Bilateral symmetry plane estimation depends on livestock segmentation results. The correct clustered results can improve the accuracy of the symmetry plane. Symmetry is found everywhere in nature. Thus, based on this feature and the definition of the CCS, the Z-axis of CCS can be obtained from the 3-D point clouds of livestock. Therefore, as in Guo et al. (2019), symmetry estimation was introduced in two phases. Firstly, the largest profile of the livestock C_* was determined. Then, a voting scheme was used to extract the symmetric plane from this largest profile, thus identifying the plane of symmetry of the livestock point cloud and also the Z-axis, which passes through the plane of symmetry.

2.2.4. *Forward direction estimation*

Through ground detection and symmetry plane estimation, the Y-axis and the Z-axis were determined. The X-axis was determined using the following equation:

$$\mathbf{n_x}=\mathbf{n_y}\times\mathbf{n_z}. \quad (2)$$

where $\mathbf{n_x}$ is the X-axis with an undetermined sign, $\mathbf{n_y}$ is the Y-axis, and $\mathbf{n_z}$ is the Z-axis with an undetermined sign. To determine the positive directions of the X-axis and Z-axis, the forward direction of the livestock must be determined. The

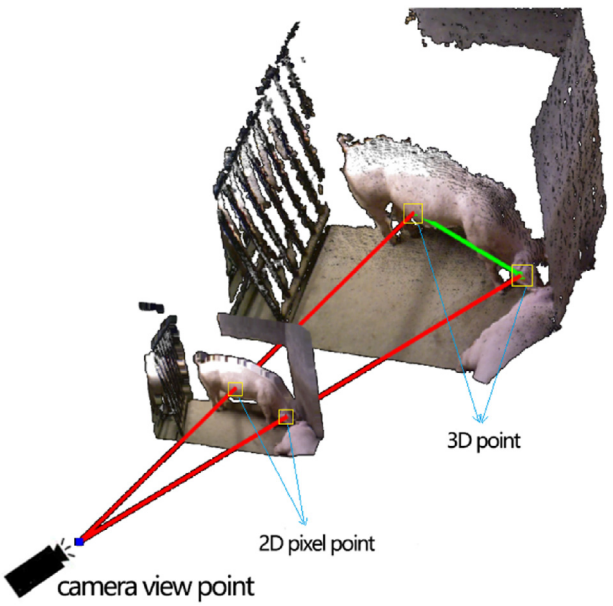


Fig. 7 – Projection of 2-D pixels into 3-D space (pig reference).

Please cite this article as: Lu, J et al., 2-D/3-D fusion-based robust pose normalisation of 3-D livestock from multiple RGB-D cameras, *Biosystems Engineering*, https://doi.org/10.1016/j.biosystemseng.2021.12.013

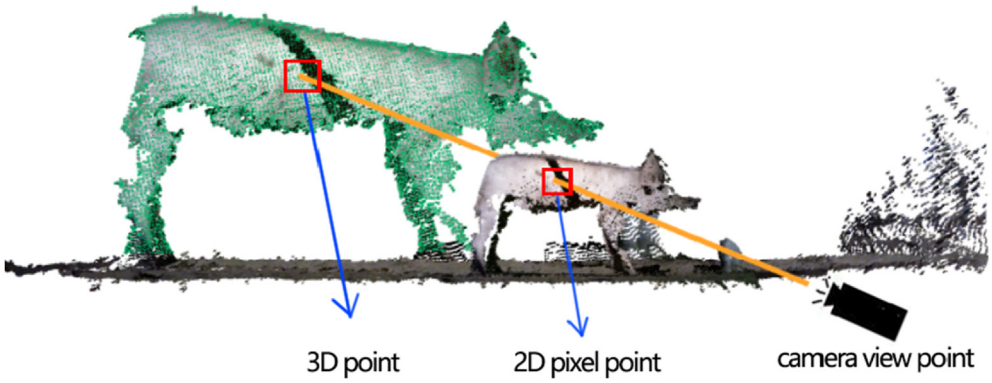


Fig. 8 – Determining the target cluster based on the 2-D pixel point (pig reference).

positive direction of the X-axis (the forward direction of the livestock) is represented by a vector. From the assumptions of Guo et al. (2019), the forward direction of the livestock can be determined. However, the front or rear part of livestock were often missing during 3-D scanning in farming conditions; thus, using this feature to judge the orientation of the livestock will frequently produce errors in forward direction estimation. Fortunately, the shortcomings of this forward direction estimation strategy can be overcome using 2-D detection technology in the proposed method.

Using the process from section 2.2.1, the targets that are represented by the centre point of the 2-D window were identified. Through the head and body or hip and body of the livestock, the forward direction of the livestock can be determined, which is represented by a vector in 2-D space. This vector can be generated by two 2-D points denoted by $\mathbf{L}_1$ and $\mathbf{L}_2$. The image use as input to predict may be an RGB image or a generated virtual image, and the generated virtual image is typically missing certain targets. Therefore, the number of targets that can be detected in different types of images is also different. Therefore, the centre point of the body region is selected as $\mathbf{L}_1$ which is easy to be detected, and selecting the centre point of the detected target region which has higher confidence in the regions of the head and hip as $\mathbf{L}_2$. Finally, among the three targets, only two (body and head, or body and hip) were used to generate the forward direction vector in 2-D space.

When the forward direction of livestock represented by a vector in 2-D space was obtained, the next step was to project the 2-D vector into 3-D space. In section 2.2.2, how to project 2-D pixels into 3-D space was introduced. Fig. 7 shows an example of the projection of 2-D pixels into 3-D space. However, when the centre point of the 2-D window was used as the head position, and the head of the livestock in the picture was not complete, the centre of the 2-D window could not fall on the pixel of the target. Fig. 9 shows a situation in which the centre of the 2-D window is not on the livestock, which can also occur during hip region detection.

To solve this problem, instead of selecting the centre point of the 2-D window of the head or hip of the livestock as $\mathbf{L}_2$, $\mathbf{L}_2$ should be generated by the relative position of the body of livestock $\mathbf{L}_1$ (x_{body}, y_{body}) and the detected target region $\mathbf{L_h}$ (x_h, y_h), which has higher confidence in the head or hip of livestock. Since the forward direction is a vector composed of two points, and the body of livestock has been determined as $\mathbf{L}_1$ as the starting point of the vector, only another point is required. However, due to the existence of the situation shown in Fig. 9, when the head or hip of the livestock is obtained through 2-D object detection, 3-D points that fall on the livestock cannot be obtained and can only determine the forward direction of livestock in 2-D. Therefore, body point $\mathbf{L}_1$ must be moved a certain distance d_l in the corresponding direction as another point $\mathbf{L}_2$ in 2-D. Then, $\mathbf{L}_2$ $(x_{direction}, y_{direction})$ can be determined for the direction vector using the following criterion:

$$\text{if } \{x_h > x_{body}\}, \text{ then } \left\{x_{direction} = x_{body} + d_l, y_{direction} = y_{body}\right\}$$

$$\text{if } \{x_h < x_{body}\}, \text{ then } \left\{x_{direction} = x_{body} - d_l, y_{direction} = y_{body}\right\}$$

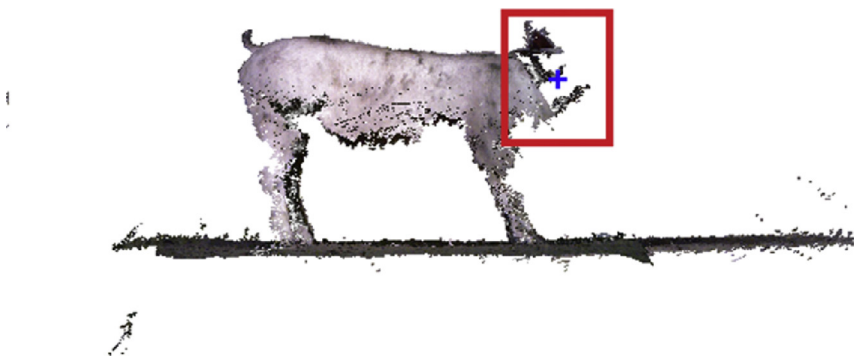


Fig. 9 – Example of when the centre of the 2-D window is not on the livestock due to missing data.

Please cite this article as: Lu, J et al., 2-D/3-D fusion-based robust pose normalisation of 3-D livestock from multiple RGB-D cameras, *Biosystems Engineering*, https://doi.org/10.1016/j.biosystemseng.2021.12.013

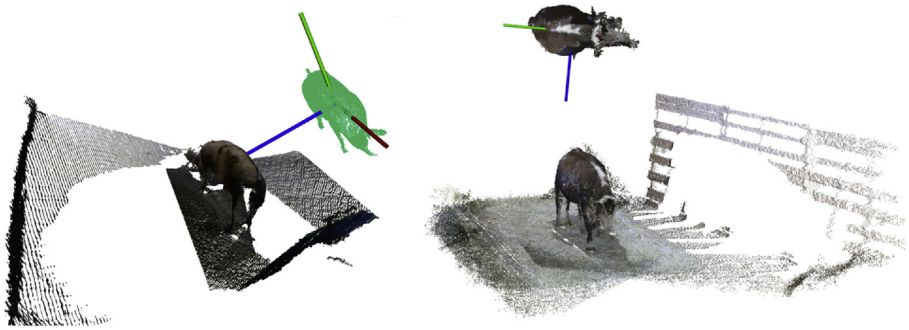


Fig. 10 – Visualisation of normalisation results for a pig (left) and a cow (right).

where d_l is the length of the direction vector and is equal to 30 pixels in this manuscript. From Eq. (1), $\mathbf{L}_1$ and $\mathbf{L}_2$ are used to generate the corresponding 3-D points denoted by $\mathbf{P}_{body}$ and $\mathbf{P}_{direction}$, respectively. Thus, the conversion from 2-D space to 3-D space is completed. Thus the forward direction vector $\mathbf{V}_f$ can be determined using the following equation:

$$\mathbf{V}_f = \mathbf{P}_{direction} - \mathbf{P}_{body} \quad 3$$

Finally, the vectors $\mathbf{n}_x$ and $\mathbf{n}_z$ must be guaranteed to serve as the X-axis and Z-axis. Therefore, their directions will be reversed when the following conditions occur:

$$\mathbf{V}_f \cdot \mathbf{n}_x < 0. \quad (4)$$

2.2.5. *Pose normalisation transformation*

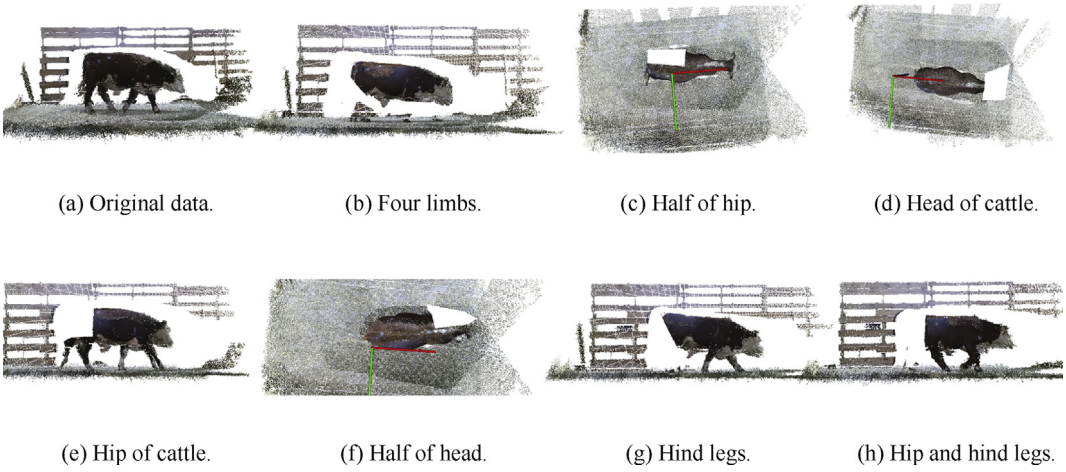
Since the X-axis, Y-axis, and Z-axis of CCS have been obtained in the previous sections, a 3-D rigid transformation can be calculated. The pose normalisation transformation can also be found in Guo et al. (2019) so only a brief summary of this procedure is given. Firstly, the centre point P_o of mass of C_* is computed using Eq. (5).

$$\mathbf{P}_o = \frac{\sum_{i=1}^{N} p_i}{N}, p_i \in C_* \quad (5)$$

where $\mathbf{P}_o$ is the centre of mass of C_* with N points. Then, the CCS with $\mathbf{n}_x$, $\mathbf{n}_y$, and $\mathbf{n}_z$ is constructed. Finally, the livestock point clouds can be placed within this coordinate system. Fig. 10 shows a visualisation of normalisation results.

2.3. *Experimental data*

Data consisting of a total of 397 point clouds was obtained by scanning in a farming environment for pigs. Point cloud data comes from Da Bei Nong Group, ShangDong, WeiHai, China which is mainly engaged in raising pigs. Each live pig was scanned from both sides, the data from each pig filtered, and finally registered. The 3-D data came from 52 selected pigs. The data was scanned frame by frame to remove inappropriate data and severely missing data. The point clouds of the scanned scene contain floors, farm facilities, and the target pigs. The spatial resolution of these point clouds was 0.005 m, which means that r is 0.005 m. In this experiment, the distance between the camera and the target was from 0.8 to 1.6 m. The second dataset included point clouds of live cattle that were gathered by other researchers (Ruchay et al., 2020), who made their data freely available to the public. They conducted RGB-D data collection on 103 cattle from three perspectives: left, right, and top. They provided the original RGB image and depth image, as well as the processed point cloud data. For more details of the farm conditions please refer to Ruchay et al. (2020). There were 309 single-view point cloud data and 103 registered point cloud data. The illuminance during data collection varied from 120 to 160 lx with an



(a) Original data. (b) Four limbs. (c) Half of hip. (d) Head of cattle.

(e) Hip of cattle. (f) Half of head. (g) Hind legs. (h) Hip and hind legs.

Fig. 11 – Examples of manually modified data and original data.

Please cite this article as: Lu, J et al., 2-D/3-D fusion-based robust pose normalisation of 3-D livestock from multiple RGB-D cameras, *Biosystems Engineering*, https://doi.org/10.1016/j.biosystemseng.2021.12.013

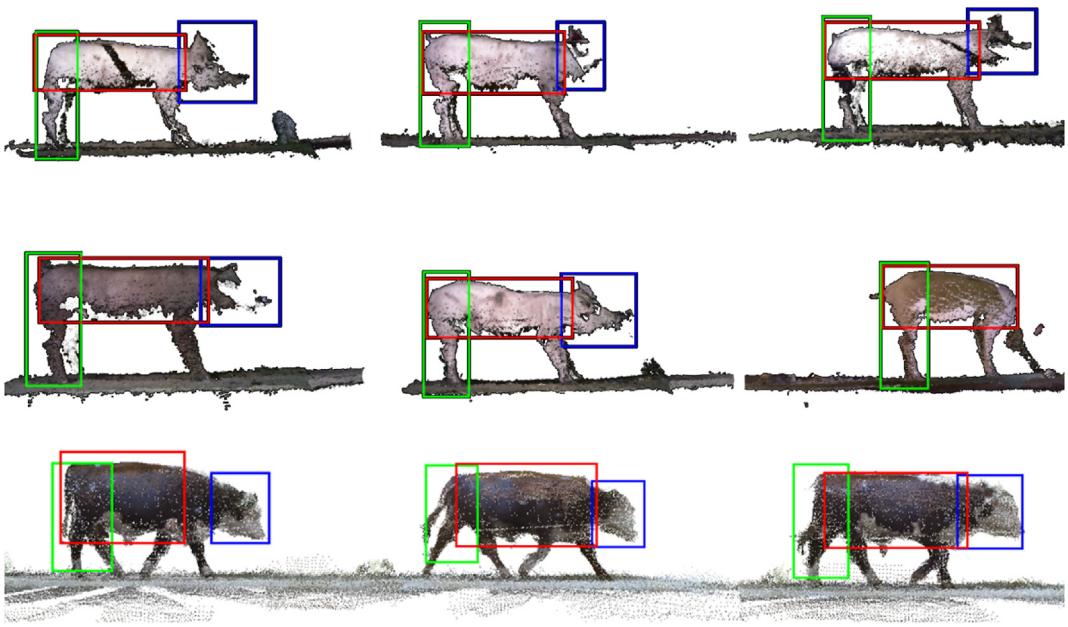


Fig. 12 – Visualisation of targets detected result (area of the body, hip, and head are labelled in red, green, and blue respectively). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

average value of 140 lx. All selected cattle were gathered at the feed fences in a separate area for image recording. The third dataset was modified from the second dataset. A dataset of 70 sets of data were randomly selected from the second dataset and the original data of the third dataset. Then, according to the deleted part, the original data was divided into on average 7 categories and modified. In Fig. 11, examples are shown corresponding to the 7 categories.

3. Results and discussion

The proposed pose normalisation method with the point clouds of livestock in a real farming environment was evaluated and the results of the proposed method compared to those of the existing method from Guo et al. (2019). The following subsections describe each experiment.

3.1. *Results of target regions detection*

In terms of target region detection, 3 categories were adopted to train and test. The initial datasets contain 1500 images, including virtual images generated from single-view RGB-D data (pig and cattle), virtual images generated from double-view RGB-D data (pig) or triple-view RGB-D data (cattle), and RGB images (pig and cattle). Additionally, for the proposed target region detector (pig and cattle), a dataset augmentation convention was used and flipped images were added to the training set. After dataset augmentation, the dataset was expanded to 4500. A total of 80% of the dataset was used for training, and 20% of the dataset was used for testing. The proposed algorithm was applied using Linux on a computer with an 8-core 2.4 GHz CPU and an RTX 2080Ti GPU. The training and testing of YOLOv4 were performed using TensorFlow, and target region detection required 0.14 s for a single image.

To evaluate the accuracy of the 2-D target detection model, the mean average precision (mAP) as the standard evaluation method of the model. When the detected box was closer to the ground-truth bounding box, the value of mAP was greater, which meant the model was more accurate.

In order to eliminate redundant scoring boxes after the initial 2-D detection, all scoring boxes overlapped with higher scoring boxes by more than 30% were removed. Similarly, all scoring boxes with very low scores (<0.05) were also removed. To calculate the mAP of the proposed detection procedure, a detection was considered to be correct if the intersection over union (IoU) of the detected target region with the ground truth box was >0.5. A ratio of the area >0.5 ensured that the forward estimation was correct. Finally, using the target region detector to predict the test set, the mAP of this detection algorithm for the proposed detection task was 94.70%. Fig. 12 shows some of the test results of the dataset.

In Fig. 12 it can be seen that the detected results of livestock body parts was more accurate and stable than those for their heads and hips, making forward direction estimation and livestock segmentation more effective. This verified that YOLOv4 is suitable for the proposed problems.

3.2. *Livestock pose normalization*

In this section, the results of the existing algorithm (SPNA) from Guo et al. (2019) were compared with the proposed algorithm. The proposed method will be denoted by OSPNA (optimised SPNA). In Table 1, the results that came from the SPNA and OSPNA are shown.

This research optimised the forward direction estimation and livestock segmentation. Therefore, these two parts of the

Please cite this article as: Lu, J et al., 2-D/3-D fusion-based robust pose normalisation of 3-D livestock from multiple RGB-D cameras, *Biosystems Engineering*, https://doi.org/10.1016/j.biosystemseng.2021.12.013

Table 1 – Results of different methods for with multi-view RGB-D data. SPNA is the existing method (Guo et al., 2019). OSPNA is the proposed method. The forward direction determination (FDD) error represents the results that the corresponding method cannot determine the forward direction of livestock. The livestock segmentation (LS) error represents the results that the corresponding method cannot determine the cluster of livestock. The pose normalisation (PN) error represents the results that the corresponding method cannot provide the correct result for pose normalisation. The supported livestock species (SLS) represents which livestock species can be supported by the corresponding method.

Method	FDD error (%)	LS error (%)	PN error (%)	SLS
OSPNA	2	2.5	4.5	Pig
SPNA	30	5	35	
OSPNA	0	2	2	Cattle
SPNA	9	2	11	

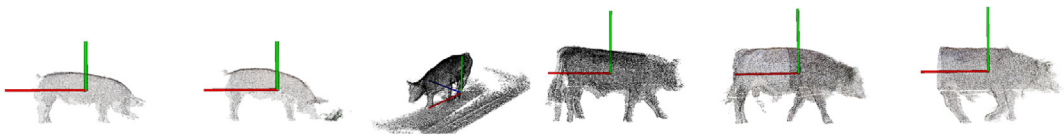
optimisation were tested with multi-view RGB-D data and compared to the results produced by of SPNA. The dataset contains 397 point clouds of live pigs and 103 point clouds of live cattle. The proposed method was shown to perform better than the existing SPNA which is the existing pose normalisation method for different species (Guo et al., 2019). As Table 1 shows quantitatively, the error of SPNA primarily comes from forward direction estimation and livestock segmentation. An incorrect forward direction estimation result will lead to the opposite direction of the pose normalisation result. In the same way, incorrect livestock segmentation results directly leads to pose normalisation failure. The proposed OSPNA optimises forward direction estimation and livestock segmentation, which substantially improves the success rate of pose normalisation. Fig. 13 shows the qualitative comparison of results produced by the two methods. The forward direction estimation and livestock segmentation in the SPNA results were incorrect in some cases. In Fig. 13 the result of the existing method (SPNA) produced different errors. As shown in Fig. 13(b), if the front or rear part of the livestock was missing, the forward direction will likely be incorrect. If the cluster of the animal contains other objects or does not describe the livestock at all, pose normalisation will fail. By introducing 2-D object detection in the proposed method, the livestock segmentation and forward direction estimation get a higher success rate, and the pose normalisation methods become more robust.

Table 2 – Results of different methods for modified cattle data. The data shown in the table shows the number of result errors in the 10 sets of data in each category.

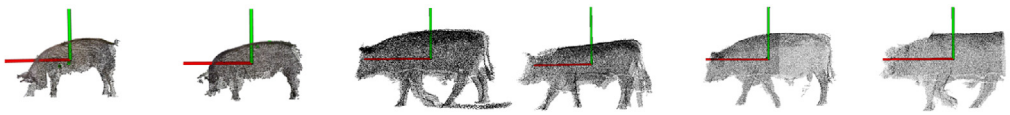
Deleted part	OSPNA error	SPNA error
Four limbs	0	2
Half of hip	2	10
Head of cattle	0	1
Hip of cattle	2	3
Half of head	0	0
Hind legs	0	0
Hip and hind legs	0	2

Through Table 1, it was found that the results of the two algorithms (SPNA and OSPNA) were quite different with pig data, whilst the difference with the cattle data was small. This was caused by the integrity of the data. From the comparison of data integrity, the integrity of cattle data was better than that of pig data. Therefore, to further verify the effectiveness of our proposed algorithm under more harsh conditions, the cattle data was manually modified. To more comprehensively verify the effectiveness of our proposed algorithm (OSPNA),

(a) Input cloud (pig and cattle)



(b) Results of SPNA (pig and cattle)



(c) Results of OSPNA (pig and cattle)

Fig. 13 – (a) show input point clouds; (b) show the results of the existing method (SPNA); and (c) show the results of the proposed pose normalisation method (OSPNA). In (b), pose normalisation failure is caused by an incorrect forward direction vector or incorrect livestock segmentation.

Please cite this article as: Lu, J et al., 2-D/3-D fusion-based robust pose normalisation of 3-D livestock from multiple RGB-D cameras, *Biosystems Engineering*, https://doi.org/10.1016/j.biosystemseng.2021.12.013

key parts of the cattle data, including the head, hind legs, limbs, hip, etc. were deleted. As shown in Figs. 11 and 7 categories were devised for the different parts to be deleted. For each category 10 sets of data were produced. Then, as shown in Table 2, the results of the two algorithms on this modified cattle data were presented. In 70 sets of data, OSPNA has 4 errors (5%) and SPNA has 18 errors (26%). Particularly in the case of half of hip deleted dataset, the results of the SPNA were almost always wrong. Examples of results are given in the last two columns of Fig. 13. Thus, after analysing the various aspects, the robustness of the OSPNA in pose normalisation was significantly improved than that of SPNA.

Through the use of multiple RGB-D cameras, large amounts of animal point cloud data can be obtained. For each animal, RGB-D data can be obtained from different angles at the same time, and the data of the entire animal can be obtained after point cloud processing. Moreover, the body shape information of animals can be obtained through pose normalisation, facilitating the continuous monitoring of the physical condition of animals. Data collected by multiple cameras can also be integrated with animal data collected by traditional manual measurements. This integrated data can also be used for research on body condition score (BCS). In farms, multiple RGB-D cameras will require more than two cameras and a computer used to synchronise the cameras. As for data storage, local storage or cloud storage can be selected. Both these storage methods have their own advantages and disadvantages and need to be selected in light of the local situation.

4. Conclusion

A 2-D/3-D fusion-based robust livestock pose normalisation method for 3-D point clouds captured from multiple RGB-D cameras, was proposed in this study. This method fused the information of 2-D images and 3-D data to improve the robustness of the computational process. Specifically, by introducing the existing 2-D object detection technology, forward estimation and segmentation for livestock were optimised. Extensive experiments with multi-view RGB-D data describing livestock showed that the proposed method is more robust and practical than existing methods. The proposed algorithm can be used in body measurement system that can be used for health monitoring, weighing, and so on. In addition, our method can deal with different livestock species, providing satisfactory input data for other algorithms, such as animal behaviour analysis and skeleton extraction.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank all those who participated in this study and provided relevant data and assistance. In addition, the author also wants to thank Da Bei Nong Group for providing us with materials for experiments. This research is supported by the National Natural Science Foundation of China [grant numbers 42071449, 41601491], The Russian Science Foundation [21-76-20014].

REFERENCES

Acharya, D., Khoshelham, K., & Winter, S. (2019). BIM-PoseNet: Indoor camera localisation using a 3D indoor model and deep learning from synthetic images. *ISPRS Journal of Photogrammetry and Remote Sensing, 150*, 245–258. https://doi.org/10.1016/j.isprsjprs.2019.02.020

Andrew, A. M. (2001). Multiple view geometry in computer vision. *Kybernetes, 30*(9/10), 1333–1341. https://doi.org/10.1108/k.2001.30.9_10.1333.2

Azzaro, G., Caccamo, M., Ferguson, J. D., Battiato, S., Farinella, G. M., Guarnera, G. C., et al. (2011). Objective estimation of body condition score by modeling cow body shape from digital images. *Journal of Dairy Science, 94*(4), 2126–2137. https://doi.org/10.3168/jds.2010-3467

Bochkovskiy, A., Wang, C., Liao, H. M., & Hao, G. (2020). *YOLOv4: Optimal Speed and Accuracy of Object Detection Point clouds processing software for livestock body measurement.* arXiv preprint arXiv:2004.10934.

Brandl, N., & Jørgensen, E. (1996). Determination of live weight of pigs from dimensions measured using image analysis. *Computers and Electronics in Agriculture, 15*(1), 57–72. https://doi.org/10.1016/0168-1699(96)00003-8

Cao, S., Du, M., Zhao, W., Hu, Y., Mo, Y., Chen, S., et al. (2020). Multi-level monitoring of three-dimensional building changes for megacities: Trajectory, morphology, and landscape. *ISPRS Journal of Photogrammetry and Remote Sensing, 167*, 54–70. https://doi.org/10.1016/j.isprsjprs.2020.06.020

Girshick, R. (2015). Fast R-CNN. In *2015 IEEE international conference on computer vision (ICCV)* (pp. 1440–1448). https://doi.org/10.1109/ICCV.2015.169

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *2014 IEEE conference on computer vision and pattern recognition* (pp. 580–587). https://doi.org/10.1109/CVPR.2014.81

Gong, Z., Lin, H., Zhang, D., Luo, Z., Zelek, J., Chen, Y., et al. (2020). A Frustum-based probabilistic framework for 3D object detection by fusion of LiDAR and camera data. *ISPRS Journal of Photogrammetry and Remote Sensing, 159*, 90–100. https://doi.org/10.1016/j.isprsjprs.2019.10.015

Guo, H., Li, Z., Ma, Q., Zhu, D., Su, W., Wang, K., et al. (2019). A bilateral symmetry based pose normalization framework applied to livestock body measurement in point clouds. *Computers and Electronics in Agriculture, 160*, 59–70. https://doi.org/10.1016/j.compag.2019.03.010

Guo, H., & Lu, J. (2021). 2D/3D fusion-based robust pose normalization of 3D livestock from multiple RGB-D cameras. Retrieved from https://gitee.com/guohaolys/robust-pose-normalization-of-3D-livestock.

Guo, H., Ma, X., Ma, Q., Wang, K., Su, W., & Zhu, D. (2017). LSSA-CAU: An interactive 3d point clouds analysis software for body measurement of livestock with similar forms of cows or pigs. *Computers and Electronics in Agriculture, 138*(Supplement C), 60–68. https://doi.org/10.1016/j.compag.2017.04.014

Hao, G., Wang, K., Su, W., Zhu, D., Liu, W., Xing, C., et al. (2017). 3D scanning of live pigs system and its application in body measurements. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,*

211–217. https://doi.org/10.5194/isprs-archives-XLII-2-W7-211-2017. XLII-2/W7.

Kawasue, K., Ikeda, T., Tokunaga, T., & Harada, H. (2013). Three-dimensional shape measurement system for black cattle using KINECT sensor. *International Journal of Circuits, Systems and Signal Processing, 7*, 222–230.

Klingenberg, C. P. (2010). Evolution and development of shape: Integrating quantitative approaches. *Nature Reviews Genetics, 11*(9), 623–635. https://doi.org/10.1038/nrg2829

Kuzuhara, Y., Kawamura, K., Yoshitoshi, R., Tamaki, T., Sugai, S., Ikegami, M., et al. (2015). A preliminarily study for predicting body weight and milk properties in lactating Holstein cows using a three-dimensional camera system. *Computers and Electronics in Agriculture, 111*, 186–193. https://doi.org/10.1016/j.compag.2014.12.020

Lahoud, J., & Ghanem, B. (2017). 2D-Driven 3D object detection in RGB-D images. In *2017 IEEE international conference on computer vision (ICCV)* (pp. 4632–4640). https://doi.org/10.1109/ICCV.2017.495

Lichti, D. D., Qi, X., & Ahmed, T. (2012). Range camera self-calibration with scattering compensation. *ISPRS Journal of Photogrammetry and Remote Sensing, 74*(Supplement C), 101–109. https://doi.org/10.1016/j.isprsjprs.2012.09.008

Li, Y., Ma, L., Tan, W., Sun, C., Cao, D., & Li, J. (2020). GRNet: Geometric relation network for 3D object detection from point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing, 165*, 43–53. https://doi.org/10.1016/j.isprsjprs.2020.05.008

Lin, T., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., et al. (2014). *Microsoft coco: Common objects in context (740-755)*. Springer. https://doi.org/10.1007/978-3-319-10602-1_48

Luhmann, T. (2010). Close range photogrammetry for industrial applications. *ISPRS Journal of Photogrammetry and Remote Sensing, 65*(6), 558–569. https://doi.org/10.1016/j.isprsjprs.2010.06.003

Luo, Z., Liu, Di, Li, J., Chen, Y., Xiao, Z., Junior, J. M., et al. (2020). Learning sequential slice representation with an attention-embedding network for 3D shape recognition and retrieval in MLS point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing, 161*, 147–163. https://doi.org/10.1016/j.isprsjprs.2020.01.003

Marchant, J. A., Schofield, C. P., & White, R. P. (1999). Pig growth and conformation monitoring using image analysis. *Animal Science, 68*(1), 141–150. https://doi.org/10.1017/S1357729800050165

Menesatti, P., Costa, C., Antonucci, F., Steri, R., Pallottino, F., & Catillo, G. (2014). A low-cost stereovision system to estimate size and weight of live sheep. *Computers and Electronics in Agriculture, 103*, 33–38. https://doi.org/10.1016/j.compag.2014.01.018

Mortensen, A. K., Lisouski, P., & Ahrendt, P. (2016). Weight prediction of broiler chickens using 3D computer vision. *Computers and Electronics in Agriculture, 123*, 319–326. https://doi.org/10.1016/j.compag.2016.03.011

Pallottino, F., Steri, R., Menesatti, P., Antonucci, F., Costa, C., Figorilli, S., et al. (2015). Comparison between manual and stereovision body traits measurements of Lipizzan horses. *Computers and Electronics in Agriculture, 118*, 408–413. https://doi.org/10.1016/j.compag.2015.09.019

Pezzuolo, A., Guarino, M., Sartori, L., González, L. A., & Marinello, F. (2018). On-barn pig weight estimation based on body measurements by a Kinect v1 depth camera. *Computers and Electronics in Agriculture, 148*, 29–36. https://doi.org/10.1016/j.compag.2018.03.003

Polewski, P., Yao, W., Heurich, M., Krzystek, P., & Stilla, U. (2017). A voting-based statistical cylinder detection framework applied to fallen tree mapping in terrestrial laser scanning point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing, 129*(Supplement C), 118–130. https://doi.org/10.1016/j.isprsjprs.2017.04.023

Qi, C. R., Liu, W., Wu, C., Su, H., & Guibas, L. J. (2018). Frustum pointnets for 3d object detection from rgb-d data. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 918–927). https://doi.org/10.1109/CVPR.2018.00102

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *2016 IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 779–788). https://doi.org/10.1109/CVPR.2016.91

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

Ruchay, A. N., Dorofeev, K. A., Kalschikov, V. V., Kolpakov, V. I., & Dzhulamanov, K. M. (2019). Accurate 3D shape recovery of live cattle with three depth cameras. *IOP Conference Series: Earth and Environmental Science, 341*(1), 12147. https://doi.org/10.1088/1755-1315/341/1/012147

Ruchay, A., Kober, V., Dorofeev, K., Kolpakov, V., & Miroshnikov, S. (2020). Accurate body measurement of live cattle using three depth cameras and non-rigid 3-D shape recovery. *Computers and Electronics in Agriculture, 179*, 105821. https://doi.org/10.1016/j.compag.2020.105821

Salau, J., Haas, J. H., Junge, W., & Thaller, G. (2017a). Automated calculation of udder depth and rear leg angle in Holstein-Friesian cows using a multi-Kinect cow scanning system. *Biosystems Engineering, 160*(Supplement C), 154–169. https://doi.org/10.1016/j.biosystemseng.2017.06.006

Salau, J., Haas, J. H., Junge, W., & Thaller, G. (2017b). A multi-Kinect cow scanning system: Calculating linear traits from manually marked recordings of Holstein-Friesian dairy cows. *Biosystems Engineering, 157*(Supplement C), 92–98. https://doi.org/10.1016/j.biosystemseng.2017.03.001

Savin, T., Kurpios, N. A., Shyer, A. E., Florescu, P., Liang, H., Mahadevan, L., et al. (2011). On the growth and form of the gut. *Nature, 476*(7358), 57–62. https://doi.org/10.1038/nature10277

Sfikas, K., Theoharis, T., & Pratikakis, I. (2011). ROSy+: 3D object pose normalization based on PCA and reflective object symmetry with application in 3D object retrieval. *International Journal of Computer Vision, 91*(3), 262–279. https://doi.org/10.1007/s11263-010-0395-x

Shen, X., Stamos, I., Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2020). Frustum VoxNet for 3D object detection from RGB-D or Depth images. In *2020 IEEE winter conference on applications of computer vision* (pp. 1698–1706). https://doi.org/10.1109/WACV45572.2020.9093276

Shuai, S., Ling, Y., Shihao, L., Haojie, Z., Xuhong, T., Caixing, L., et al. (2020). Research on 3D surface reconstruction and body size measurement of pigs based on multi-view RGB-D cameras. *Computers and Electronics in Agriculture, 175*, 105543.

Song, X., Bokkers, E. A. M., van der Tol, P. P. J., Koerkamp, P. W. G. G., & van Mourik, S. (2018). Automated body weight prediction of dairy cows using 3-dimensional vision. *Journal of Dairy Science, 101*(5), 4448–4459. https://doi.org/10.3168/jds.2017-13094

Tasdemir, S., Urkmez, A., & Inal, S. (2011). Determination of body measurements on the Holstein cows using digital image analysis and estimation of live weight with regression analysis. *Computers and Electronics in Agriculture, 76*(2), 189–197. https://doi.org/10.1016/j.compag.2011.02.001

Tassinari, P., Bovo, M., Benni, S., Franzoni, S., Poggi, M., Mammi, L. M. E., et al. (2021). A computer vision approach based on deep learning for the detection of dairy cows in free stall barn. *Computers and Electronics in Agriculture, 182*, 106030. https://doi.org/10.1016/j.compag.2021.106030

Thomson, C., & Boehm, J. (2015). Automatic geometry generation from point clouds for BIM. *Remote Sensing, 7*(9), 11753–11775. https://doi.org/10.3390/rs70911753

Please cite this article as: Lu, J et al., 2-D/3-D fusion-based robust pose normalisation of 3-D livestock from multiple RGB-D cameras, *Biosystems Engineering*, https://doi.org/10.1016/j.biosystemseng.2021.12.013

Viazzi, S., Bahr, C., Van Hertem, T., Schlageter-Tello, A., Romanini, C. E. B., Halachmi, I., et al. (2014). Comparison of a three-dimensional and two-dimensional camera system for automated measurement of back posture in dairy cows. *Computers and Electronics in Agriculture, 100*, 139–147. https://doi.org/10.1016/j.compag.2013.11.005

Vosselman, G., Coenen, M., & Rottensteiner, F. (2017). Contextual segment-based classification of airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing, 128*(Supplement C), 354–371. https://doi.org/10.1016/j.isprsjprs.2017.03.010

Wang, K., Guo, H., Ma, Q., Su, W., Chen, L., & Zhu, D. (2018). A portable and automatic Xtion-based measurement system for pig body size. *Computers and Electronics in Agriculture, 148*, 291–298. https://doi.org/10.1016/j.compag.2018.03.018

Wang, R., Hu, Y., Wu, H., & Wang, J. (2016). Automatic extraction of building boundaries using aerial LiDAR data. *Journal of Applied Remote Sensing, 10*, 10. https://doi.org/10.1117/1.JRS.10.016022

Wongsriworaphon, A., Arnonkijpanich, B., & Pathumnakul, S. (2015). An approach based on digital image analysis to estimate the live weights of pigs in farm environments. *Computers and Electronics in Agriculture, 115*, 26–33. https://doi.org/10.1016/j.compag.2015.05.004

Xiang, X., Wang, Z., Lao, S., & Zhang, B. (2020). Pruning multi-view stereo net for efficient 3D reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing, 168*, 17–27. https://doi.org/10.1016/j.isprsjprs.2020.06.018

Yang, B., Liu, Y., Dong, Z., Liang, F., Li, B., & Peng, X. (2017). 3D local feature BKD to extract road information from mobile laser scanning point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing, 130*(Supplement C), 329–343. https://doi.org/10.1016/j.isprsjprs.2017.06.007

Yilmaz, O., Cemal, I., & Karaca, O. (2013). Estimation of mature live weight using some body measurements in Karya sheep. *Tropical Animal Health and Production, 45*(2), 397–403.

Zhang, Y., Shen, B., Wang, S., Kong, D., & Yin, B. (2018). L0-regularization-based skeleton optimization from consecutive point sets of kinetic human body. *ISPRS Journal of Photogrammetry and Remote Sensing, 143*, 124–133. https://doi.org/10.1016/j.isprsjprs.2018.04.016

Please cite this article as: Lu, J et al., 2-D/3-D fusion-based robust pose normalisation of 3-D livestock from multiple RGB-D cameras, *Biosystems Engineering*, https://doi.org/10.1016/j.biosystemseng.2021.12.013