Object detection and tracking using a high-performance artificial intelligence-based 3D depth camera: towards early detection of African swine fever

Harry Wooseuk Ryu 1,2, Joo Ho Tai 1,*

1Research Institute for Veterinary Science, College of Veterinary Medicine, Seoul National University, Seoul 08826, Korea
2Department of Computer Science, University of Toronto, Toronto, Ontario, M5S 1A1, Canada

ABSTRACT

Background: Inspection of livestock farms using surveillance cameras is emerging as a means of early detection of transboundary animal disease such as African swine fever (ASF). Object tracking, a developing technology derived from object detection aims to the consistent identification of individual objects in farms.

Objectives: This study was conducted as a preliminary investigation for practical application to livestock farms. With the use of a high-performance artificial intelligence (AI)-based 3D depth camera, the aim is to establish a pathway for utilizing AI models to perform advanced object tracking.

Methods: Multiple crossovers by two humans will be simulated to investigate the potential of object tracking. Inspection of consistent identification will be the evidence of object tracking after crossing over. Two AI models, a fast model and an accurate model, were tested and compared with regard to their object tracking performance in 3D. Finally, the recording of pig pen was also processed with aforementioned AI model to test the possibility of 3D object detection.

Results: Both AI successfully processed and provided a 3D bounding box, identification number, and distance away from camera for each individual human. The accurate detection model had better evidence than the fast detection model on 3D object tracking and showed the potential application onto pigs as a livestock.

Conclusions: Preparing a custom dataset to train AI models in an appropriate farm is required for proper 3D object detection to operate object tracking for pigs at an ideal level. This will allow the farm to smoothly transit traditional methods to ASF-preventing precision livestock farming.

Keywords: 3D depth camera; object detection; artificial intelligence; African swine fever; precision livestock farming
INTRODUCTION

Transboundary animal diseases (TADs) are highly contagious and transmissible epidemic diseases, with the potential to spread rapidly [1]. Among the TADs, African swine fever (ASF) has high mortality rates in domestic pigs and wild boar that are nearing 100%, which is causing major losses in the swine industry worldwide [2,3]. South Korea’s pig industry holds the 2nd largest share of the agricultural sector, topped only by rice. Here, the disease is still causing great concern, with transmission from ASF-infected wild boars to domestic pigs highly probable [4]. The first outbreak of ASF in South Korea was reported in September 2019, following an outbreak in North Korea, as well as dramatic spread in China and other Asian countries [5]. ASF has been reported in at least 60 countries around the globe [1]. Currently, no approved vaccine or treatment for ASF is available. For this reason, early detection of TADs such as ASF in the field followed by rapid laboratory diagnosis is critical to its prompt eradication. Passive surveillance based on reports from farmers is highly recommended for the early detection of ASF, ensuring the effective prevention of further spread and significant minimization of insurmountable damages as well as huge socio-economic consequences in the country [1,3,5,6].

Considering that the average farm worker in the pig industry is responsible for approximately 2,000 pigs, it is unfeasible for such a small number of farm workers to provide the proper care to individual animals in medium to large-scale livestock farms [7]. Recently, the use of surveillance cameras for the inspection of livestock has been emerging as a replacement for traditional methods to reduce labor in farms [8]. With artificial intelligence (AI)-based technology rapidly integrating cameras, it is now gradually being applied to the field of veterinary medicine, bringing about the emergence of disease-preventing or disease-free precision livestock farming (PLF) [9-11]. These camera systems enable farms to monitor not only the normal behaviors of livestock such as movement, relaxation, eating, drinking, and sleeping but also abnormal behaviors, i.e. aggression or illness showing various clinical symptoms occurring due to the TAD infection [6,12].

The underlying core technology in such systems is object detection, which allows computer systems to detect objects in the same manner as humans. Object tracking is an extended technology derived from object detection, in which “the target is specified in the first frame and must be detected and tracked in the next frames of the video” [13]. The key in object tracking is to accumulate consecutive data of the target. This accumulated data may be the target’s coordinates, direction, velocity or total distance moved. Today, the analysis of such data through the use of AI is widely performed in a range of applicable areas, including but not limited to computer science, safety monitoring systems, intelligent transportation, hospitals, sports [14] or even farms.

This study was conducted to test whether a high-performance AI-based 3D depth camera can accurately detect objects such as individual humans, and, preliminarily, individual pigs using a 3D bounding box, before its practical application. In addition, some degree of object tracking is examined to determine whether the use of object detection and 3D depth cameras is viable in livestock tracking.
MATERIALS AND METHODS

3D Depth sensor and other hardware components
A high-performance AI-based 3D depth camera (ZED 2, Stereolabs Inc., USA) was used, and the acquired 3D depth data was processed on a laptop.

The hardware specifications of the laptop used was an Intel Core i5-4310M running at 2.7 GHz, 8GB RAM, and NVIDIA GeForce 840m.

Generally, ZED 2 is better than other accessible 3D depth cameras in terms of the parameters shown in the following comparison of depth cameras (Table 1). It can cover a wider range with higher resolution, which is evident in the depth field of view (FOV) and depth resolution. Although the depth frames per second (fps) varies with depth resolution, ZED 2 has the highest available depth resolution and fps. It also has the longest ideal range for depth, which is 19.7 m in total.

Software
DirectX, Unity and two AI models were integrated on the hardware above. The several object detection models adopted from the object detection modules of 3D depth camera co., ZED software development kit (SDK) were a fast detection model and an accurate detection model for human body and/or multi class box (animals) for pig (www.stereolabs.com). The fast detection model had shorter processing time, but lower accuracy; the accurate detection model had longer processing time and higher accuracy. To exploit the 3D depth video data, a 3D bounding box technique was integrated into both models.

3D bounding box
There is a significant difference between a 2D bounding box and a 3D bounding box. Adding another dimension to bounding boxes allows further analysis such as size, orientation or even posture of a given object. For example, a 2D bounding box only has to locate and fit a square around the given object. But for 3D bounding boxes, the size as well as the depth should be estimated so that the software can fit a cube around the given object. Most software which aims to produce a 3D bounding box starts with a 2D bounding box prediction. Then, the algorithm identifies the key points within and around the 2D bounding box through machine learning such as convolutional neural network (CNN) or regional(R)-CNN [15], and these are then used to create the final output of a 3D bounding box. Currently, 3D bounding boxes are frequently used in the fields of robotics and augmented reality, but soon will be extended and readily used in various other fields.

Table 1. Comparison of specs between ZED 2 and other 3D depth cameras

| Variables             | ZED 2 (2020) | Kinect v1 (2010) | Kinect v2 (2014) | D415 (2018) | D435 (2018) |
|-----------------------|--------------|------------------|------------------|-------------|-------------|
| Depth FOV (H × V)     | 110° × 70°   | 57° × 43°        | 70.6° × 60°      | 65° × 40°   | 87° × 58°   |
| Depth resolution      | Up to 4,416 × 1,242 | 320 × 240     | 512 × 424        | Up to 1,280 × 720 | Up to 1,280 × 720 |
| Depth fps             | Up to 100 fps | Up to 30 fps    | Up to 30 fps     | Up to 90 fps | Up to 90 fps |
| Ideal range for depth | 0.3–20 m    | 0.4–4.5 m       | 0.5–4.5 m        | 0.5–3 m     | 0.3–3 m     |
| RGB FOV (H × V)       | 110° × 70°   | 62° × 48.6°     | 84.1° × 53.8°    | 69° × 42°   | 69° × 42°   |
| RGB resolution and fps| Up to 4,416 × 1,242 | Up to 1,280 × 720 | 1,920 × 1,080 at 30 fps | 1,920 × 1,080 at 30 fps | 1,920 × 1,080 at 30 fps |

FOV, field of view; H × V, horizontal × vertical; fps, frame per second; RGB, red, green and blue.
Data collection and processing

While recording a scene with an AI-based 3D depth camera, multiple crossovers were simulated by two human beings to seek any potential of object tracking. The direction of crossing was perpendicular to the camera's view.

The crossing was started in a stand-still position, in which each person is to the right and left side of the camera. While both persons walk towards each other simultaneously, the first crossing aligns the camera with the first person and then the second person. After the crossing, both persons stand still for 5 seconds. Then the second crossing aligns the camera with the second person, and the first person. Both persons stand still for 5 seconds, and the procedure is repeated.

After collecting multiple crossovers, an inspection was done for each crossover. This was to seek any potential for object tracking by looking at the change of the identification number (ID) and 3D bounding boxes.

As a preliminary test for the application to a livestock farm, ZED 2 was installed in an experimental pig pen, 2.5 m above the center of the pen and with a field of view covering the entire pen area (Fig. 1). For this reason, the camera showed a top view. The acquired photo/video was processed using the aforementioned AI model to determine whether it has potential for object detection of pigs in 3D, implying the future possibility of development and usage for the early detection of TADs.

RESULTS

Fig. 2 shows the two different AI models implemented: the fast detection model and the accurate detection model. A frame was taken out of the video to visualize the crossing as well as the object detection. ‘Before’ indicates the frame where the two persons stood still to the left and right sides of the camera. ‘After’ indicates the frame after the crossing happened, and the two persons stood still. ‘Crossing’ indicates the frame just before the crossing happens.

The 3D depth camera has successfully identified human beings. Both the fast and the accurate model were capable of detecting the human body (Fig. 2). Every detection by the two models
had a 3D bounding box, ID, and estimated distance from camera. In addition, the color of the text (ID and distance) as well as the 3D bounding box were consistent for each object.

Both models had the same ID value for the person who was in front while crossing. That is to say, both models had a consistent ID and color for the person who was closer to the camera (and therefore, was seen at all times). In the fast detection model in Fig. 2, a person with the ID of 0 (green) remained the same throughout (A), (B), and (C). Similarly, in the accurate detection model, a person with the ID of 1 (blue) remained the same throughout (D), (E), and (F).

The result for the person who was behind while crossing, and therefore was unseen by the camera, was different for each model. The fast detection model failed to provide the same ID for the same object. (A) and (C) have different ID numbers and colors: 5 (blue) and 8 (red) (Fig. 2 upper row). On the other hand, the accurate detection model showed consistency in providing the same ID. (D) and (F) have the same number as IDs, and the same colored 3D bounding box for both persons (Fig. 2, lower row). Therefore, the accurate detection model showed potential for object tracking.

When object detection is prepared for livestock farms, farms can track individual livestock through the integration of a 3D depth camera. Fig. 3 shows how the technology is used when object detection is integrated into livestock farms. The object tracking technology is applied

Fig. 3. Illustration of future potential for livestock detection (here, pig as livestock object). The object detection technology is applied to achieve individual object detection. Both (A) and (B) show what it might look like if this technology were properly applied to data acquired from a pig pen. A pig was detected and designated as ID: 25 and estimated distance of 2.51 m from the 3D depth camera (A). The detection example of two pigs were shown in (B), but different ID and estimated distance from the camera (5 and 2.64, respectively) as well as different color and shape in the object detection box and text from those in (A) were provided.

ID, identification number.
so that each individual livestock is detected and tracked. Both (A) and (B) are the result of applying the existing object detection model to pigs, showing a box around each pig or pigs with an ID and their distance from the camera. **Fig. 3A** shows that a pig was detected by the object detection model, providing an ID and estimated distance from camera. The model also detected an object of two pigs (**Fig. 3B**), while its ID and estimated distance from camera as well as color and shape in the box and text provided were all different from those in **Fig. 3A**.

**DISCUSSION**

We have used ZED 2, a recently developed high-performance AI-based 3D depth camera, to seek the potential of livestock tracking.

Similar to depth, ZED 2 also shows better performance in terms of RGB. It has greater range, with highest available RGB resolution and fps (**Table 1**). Due to higher FOV in both depth and RGB (**Table 1**), installation of ZED 2 can also resolve the issue of having to install multiple cameras and perform data-intensive tasks such as co-registering the images (**Fig. 1**).

A sequence of images is called a video and each individual image within a video is called a frame. A video that has 15 images within one second is said to have 15 fps. If the input is an image, the object is detected with a non-moving bounding box around it and labelled on the output image. If the input is a video, however, the output would have a moving bounding box around the object. But in essence, the object detector is processing each image and producing an output of each image. The output images are then sequenced so that it looks like the bounding box is moving. Therefore, the outputs for both object detection and tracking may look the same.

Mcfarlane and Schofield reported the loss of tracking due to large, unpredictable movements by the piglets as a limitation of object tracking [16]. A solution to this problem is to implement a camera with higher fps. The ZED 2, which allows the highest fps, provides the best resolution among comparable products (**Table 1**) for the ‘sudden movement’ problem in object tracking.

On one hand, object detection allows computers to answer the ‘what’ question for objects in a given image. However, object detection falls short when it comes to answering the ‘which’ question. It is generally incapable of differentiating individual objects of the same classification. For example, object detection can ‘detect’ dogs if it is trained to detect dogs by the AI. Nonetheless, it cannot individually ‘identify’ every single dog, which means it cannot answer the question of ‘which’ dog moved the most, for example.

For this reason, diverse explorations and investigations must be performed on various kinds of individual farms. This includes different sizes, settings, and environments of farm management systems. To the best of the author’s knowledge, object tracking, especially in 3D, is merely conducted in comparison to object detection (ratio-wise), in the fields of veterinary medicine and livestock industries or even in computer science. Thus, in order to effectively fuse these advanced sensor and AI-based technologies into the veterinary and livestock areas, a preliminary investigation was conducted. An analysis of both the benefits and the pitfalls or challenges is essential to reduce trial-and-error, and to better utilize such advanced technology.
In the study results shown in Fig. 2, two human beings (the two authors of this paper) simulated livestock crossing over, being recorded in front of the 3D depth camera as a side view (direction of crossing perpendicular to the camera’s view). In this setting, the person who was front (closer to the camera) could completely conceal the person behind in all multiple crossing-overs. Our results have shown that both the ‘fast’ and the ‘accurate’ AI models were able to provide 3D bounding boxes, ID, and distance measurement on two human beings, and the accurate AI model performed object tracking better than the faster AI model, demonstrating that the accurate object detection model has greater potential.

As such, although the final goal of our study application focuses primarily on livestock tracking, human detection was also implemented and tested in livestock farms. Detection and tracking of farm staff/workers, guests or intruders/outsiders may offer early evidence in active surveillance for ASF at farms [17]. Greater attention to TADs which are both zoonotic and non-zoonotic (i.e. ASF, Newcastle disease, and Peste des petits ruminants) should be fully considered within the framework of One Health [18].

In addition to a general 2D sensor camera, a 3D depth sensor which produces depth image data consisting of the distance values from the sensor for each pixel allows us to “better understand the images in 2D and extend the image-level knowledge to the real world” [15]. The integration of depth image data into generic RGB image data overcomes the “disadvantages of 2D computer vision” [19] and “increases (the) accuracy and robustness” [20] of object tracking. Thus, an extensive analysis of depth information will allow for the further enhancement of the performance of object detection or object tracking, while overcoming some of the challenges mentioned above.

Depth sensors are not always beneficial. According to a study by Mithun et al. [21], depth sensors bring additional challenges including ‘retraining already well trained RGB based AI,’ ‘low accessibility and limited availability of depth images’ and ‘demands costly hardware.’ These issues remain common around the utilization of depth information. Such problems should be either resolved or avoided by discovering a new pathway to allow farms to integrate object detection and tracking effectively.

In relation to real-time processing, a paper from HITEC university recommends how to approach this issue [22]. Using only software solutions for complex object tracking is not viable due to limitations in processing capabilities. They also claim that to use object tracking in real time, finding a system which is fast, power efficient, and memory manageable is essential. Thus, both software and hardware must be cohesive and optimized in order to utilize advanced object detection and tracking.

Since there are multiple livestock recorded on camera, the idea of multiple object tracking should be feasible. Multiple Object Tracking aims to locate multiple objects, maintaining their identities, and accumulate the movement data for each object [23]. This requires further optimization and management of both hardware and software in order to minimize the capital burden on farms, and for it to be implemented in farms’ management systems.

The preliminary result data in Fig. 3 show both the strengths and weaknesses of the use of 3D bounding boxes. Fig. 3A shows a pig within a 3D bounding box, which is evidence of successful detection. Such detections will allow further analysis of data including movement or behavioral changes of the pig to determine whether a TAD infection should be suspected.
or not. Fig. 3B shows the image of two pigs within a 3D bounding box. Since the aim of object tracking is to track individual objects, the current detection AI model falls short when it comes to detecting pigs, given a scenario of multiple pigs in a pen. In this study, ZED 2 was used as a tool for detecting/tracking objects in a limited range of object subclasses and it produced several types of data as shown in our result figures. Thus, it is necessary both to prepare and train the pig data set to derive a suitable AI model according to the experimental or field environment to enhance the detection as well as the object tracking of the pig for the future early detection of clinical signs of TADs, such as ASF infection.

As wildlife, wild boar and ASFv-positive wild boar carcasses are increasing concerns of the risk of spillover of ASFv between wild boar to domestic pigs [4,24,25]. The detection and identification of wild boars using the 3D depth camera and AI model seems also to be very useful technology. However, the wild boar is quite different from the domestic pig in its exterior appearance, such as fur color and hair quantity, teeth, head, neck, trunk, etc [26]. Furthermore, the wild boar has a large and diverse spectrum of habitat environments, compared to the static environment of backyard or indoor farmed pigs. For these reasons, the object detection of a wild boar, like other wildlife, involves lots of complexity and trial and error from the initial step of the collection and processing of its large data set in order to extract its characteristic features from various camouflaging components of environmental backgrounds as well [27]. Therefore, to apply an AI-based 3D depth camera to the development of object detection of wildlife such as wild boar and generate appropriate result data (ex. number, ID in a confined area, staying duration, emerging time and so on) for further analysis, a large data set for wild boars, plus a great deal of time and effort, are required prior to the development of software or algorithm, through cooperation with the authorities concerned such as the wildlife rescue and management centers. Once developed and working cohesively well with the hardware (ex. ZED 2), such convergence methods and research direction will certainly offer new possibilities in veterinary and environmental biosecurity to stop the spread of ASF.

A robust and accurate object detection model should be implemented and well-trained in order to successfully track objects in terms of the challenges and goals for future research and improvement. With such a detection model, the most suitable hardware should be chosen to fulfill the demands on processing power. Finally, an efficient method to handle the 3D depth video data must be implemented to reduce the burdens on farms.

In conclusion, this study encourages researchers and livestock farms to apply an accurate and well-trained AI model that can serve as a core of the future smart farm or PLF. Preparing a custom data set in an appropriate model farm to train AI models for proper 3D object detection will be also critically important in operating object tracking at an ideal level, which is insufficient for application to pigs. This will enable farms to detect the signs and symptoms of ASF as early as possible in a contact-free and automatic way, and move towards a smooth transition to precision TAD-preventing livestock farming in the near future.

**ACKNOWLEDGEMENTS**

We would also like to thank our team members at the Research Institute for Veterinary Science in the College of Veterinary Medicine of Seoul National University and Foreign Animal Disease Division, Animal and Plant Quarantine Agency, Gimcheon, Korea for their helps for this study.
REFERENCES

1. Clemmons EA, Alfonso KJ, Dutton JW 3rd. Transboundary animal diseases, an overview of 17 diseases with potential for global spread and serious consequences. Animals (Basel). 2021;11(7):2039.

2. Dixon LK, Sun H, Roberts H. African swine fever. Antiviral Res. 2019;165:34-41.

3. Faverjon C, Meyer A, Howden K, Long K, Peters L, Cameron A. Risk-based early detection system of African swine fever using mortality thresholds. Transbound Emerg Dis. 2021;68(3):1151-1161.

4. Chenais E, Stähli K, Gaberti V, Depner K. Identification of wild boar-habitat epidemiologic cycle in African swine fever epizootic. Emerg Infect Dis. 2018;24(4):810-812.

5. Cho KH, Kim HJ, Kim DY, Yoo D, Nah JJ, Kim YJ, et al. Surveillance of ASF-infected pig farms from September to October 2019 in South Korea. J Vet Sci. 2021;22(2):e26.

6. Yoon H, Hong SK, Lee I, Yoo DS, Jung CS, Lee E, et al. Clinical symptoms of African swine fever in domestic pig farms in the Republic of Korea, 2019. Transbound Emerg Dis. 2020;67:2245-2248.

7. Kim J, Chung Y, Choi Y, Sa I, Kim H, Chung Y, et al. Depth-based detection of standing-pigs in moving noise environments. Sensors (Basel). 2017;17(12):2757.

8. Carr J, Chen SP, Conner JF, Kirkwood R, Segalès J. Pig Health. Boca Raton: CRC Press; 2018.

9. Franken E, Berckmans D. Precision livestock farming for pigs. Anim Front. 2017;7(1):32-37.

10. Queen C. Artificial Intelligence in Veterinary Medicine [Internet]. Horsham: Al Med; https://ai-med.io/ai-med-news/artificial-intelligence-veterinary/. Updated 2018.

11. Cihan P, Gököz E, Kalipsiz O. A review of machine learning applications in veterinary field. Kafkas Univ Vet Fak Derg. 2017;23(4):673-680.

12. Aiello SE, Moses MA, Allen DG. The Merck Veterinary Manual. 11th ed. Kenilworth, NJ: Merck Sharp & Dohme Corp; 2016.

13. Soleimanitaleb Z, Keyvanrad MA, Jafari A. Proceedings of 2019 9th International Conference on Computer and Knowledge Engineering (ICCKE); 2019 Oct 24–25; Mashhad, Iran. Mashhad: Ferdowsi University of Mashhad; 2019.

14. Bi F, Ma X, Chen W, Fang W, Chen H, Li J, et al. Review on video object tracking based on deep learning. J New Media. 2019;1(2):63-74.

15. Zhao ZQ, Zheng P, Xu ST, Wu X. Object detection with deep learning: a review. IEEE Trans Neural Netw Learn Syst. 2019;30(11):3212-3232.

16. McFarlane NJ, Schofield CP. Segmentation and tracking of piglets in images. Mach Vis Appl. 1995;8(3):187-193.

17. Cho KH, Kim HJ, Kim DY, Yoo D, Nah JJ, Kim YJ, et al. Surveillance of ASF-infected pig farms from September to October 2019 in South Korea. J Vet Sci. 2021;22(2):e26.

18. Torres-Velez E, Havas KA, Spiegel K, Brown C. Transboundary animal diseases as re-emerging threats - impact on one health. Srman Diagn Pathol. 2019;36(3):193-196.

19. Chen I, Chi C, Hsu S, Chen L. Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2014 May 4–9; Florence, Italy. Piscataway: IEEE; 2014.

20. Gutev A, Debono CJ. Proceeding of the IEEE EUROCON 2019 - 18th International Conference on Smart Technologies; 2019 Jul 1–4; Novi Sad, Serbia. Piscataway: IEEE; 2019.

21. Mithun NC, Munir S, Guo K, Shelton C. Proceedings of the 2018 17th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN); 2018 Apr 11-13; Porto, Portugal. Piscataway: IEEE; 2018.

22. Naeem H, Ahmad J, Tayyab M. Proceedings of the INMIC; 2013 Dec 19–20; Lahore, Pakistan. Piscataway: IEEE; 2013.

https://doi.org/10.4142/jvs.21252
23. Luo W, Xing J, Milan A, Zhang X, Liu W, Zhao X, et al. Multiple object tracking: a literature review. Artif Intell. 2021;203:103448.
24. Dixon LK, Stahl K, Jori F, Vial L, Pfeiffer DU. African swine fever epidemiology and control. Annu Rev Anim Biosci. 2020;8(1):221-246. PUBMED | CROSSREF
25. Lim JS, Kim E, Ryu PD, Pak SI. Basic reproduction number of African swine fever in wild boars (Sus scrofa) and its spatiotemporal heterogeneity in South Korea. J Vet Sci. 2021;22(5):e71. PUBMED | CROSSREF
26. Tack J. Wild Boar (Sus scrofa) Populations in Europe: A Scientific Review of Population Trends and Implications for Management. Brussels: ELO; 2018, 56.
27. Mondal A. Camouflaged object detection and tracking. arXiv. 2020 Dec 25. https://arxiv.org/abs/2012.13581.