

Deep Learning-based Overlapping-Pig Separation by Balancing Accuracy and Execution Time

Hanhaesol Lee Dept. of Computer Convergence Software Korea University Korea 30019, Sejong maxsoribada@korea.ac.kr	Jaewon Sa Dept. of Computer Convergence Software Korea University Korea 30019, Sejong sjwon92@korea.ac.kr	Yongwha Chung Dept. of Computer Convergence Software Korea University Korea 30019, Sejong ychungy@korea.ac.kr
Daihee Park Dept. of Computer Convergence Software Korea University Korea 30019, Sejong dhpark@korea.ac.kr	Hakjae Kim Class Act, Co., Ltd. Digital-ro, Geumcheon-gu Korea 08589, Seoul krunivs@gmail.com	

ABSTRACT

The crowded environment of a pig farm is highly vulnerable to the spread of infectious diseases such as foot-and-mouth disease, and studies have been conducted to automatically analyze behavior of pigs in a crowded pig farm through a video surveillance system using a top-view camera. Although it is required to correctly separate overlapping-pigs for tracking each individual pigs, extracting the boundaries of each pig fast and accurately is a challenging issue due to the complicated occlusion patterns such as X shape and T shape. In this study, we propose a fast and accurate method to separate overlapping-pigs not only by exploiting the advantage (i.e., one of the *fast* deep learning-based object detectors) of You Only Look Once, YOLO, but also by overcoming the disadvantage (i.e., the *axis aligned* bounding box-based object detector) of YOLO with the test-time data augmentation of rotation. Experimental results with the occlusion patterns between the overlapping-pigs show that the proposed method can provide better accuracy and faster processing speed than one of the state-of-the-art deep learning-based segmentation techniques such as Mask R-CNN (i.e., the performance improvement over Mask R-CNN was about 11 times, in terms of the accuracy/processing speed performance metrics).

Keywords

Pig monitoring, overlapping-pigs, separation, deep learning, YOLO.

1. INTRODUCTION

The early detection of management problems related to health and welfare is an important aspect of caring for group-housed livestock. In particular, caring for individual animals is necessary to minimize the possible damage caused by infectious diseases or other health and welfare problems. However, it is impossible to carefully manage pigs because each farm worker in Korea has to manage 1,800 pigs on the average. In order to solve these problems, various researches have been reported to automatically manage the behavior of individual pigs using surveillance cameras [1-5].

For surveillance systems using camera, separating dense pigs into individual pigs in a crowded

environment is an essential element in the behavioral analysis of individual pigs. Recently, a faster deep learning-based object detector, such as You Only Look Once (YOLO) [6], has been developed, and thus method of the results of pig detection by using YOLO have been reported. Furthermore, separating of touching-pigs into individual pigs in crowded environments have been reported by using YOLO [7, 8]. Since YOLO has some limitation in detecting small objects, however, separating overlapping-pigs with YOLO is challenging and has not been reported yet. In complicated occlusion patterns such as X shape and T shape, for example, the small areas of the occluded pig may not be detected accurately with YOLO only.

In this study, we propose a separation method for overlapping-pigs by considering accuracy and

execution time. Our proposed method consists of the foreground detection step as the preprocessing and separation step for the overlapping-pigs. In the first step as the preprocessing, we consider data fusion between the infrared and depth information which is obtained from an Intel RealSense camera [9]. Because the infrared information is more accurate than the depth information, the infrared information can compensate for the defect of the depth information such as inaccurate pixels. Also, the depth information is less affected by illumination conditions, so that it can be complementary with the infrared information for precisely detecting pigs. Second, simple and effective image processing techniques are exploited for fast detection of pigs to satisfy the real-time execution.

In the separation of overlapping-pigs steps, we propose a detection and separation method of overlapping-pigs using YOLO and image processing techniques. First, the test image is augmented by rotating it at various angles such that the YOLO boundary box contains the occluding pig only, regardless of the occlusion pattern of the test image. In particular, we use a pre-computed lookup table for the rotation in real-time. Then, by performing YOLO on the augmented data including the input data, a bounding box for each rotation is obtained, and the optimal angle of the rotation is determined. Finally, we separate the overlapping-pigs by using the YOLO result and some image processing techniques.

The rest of the paper is structured as follows: we describe background in Section 2, and Section 3 explains the proposed method for detecting pigs from the various factors in the pig room and for separating overlapping-pigs from the various occlusion shapes. Section 4 shows the experimental results of the proposed method, and the paper is finally concluded in Section 5.

2. BACKGROUND

Recently, a variety of methods have been proposed for separating touching objects using images acquired from a camera. As an example, a separation method for two touching pigs has been proposed by analyzing the outline of touching pigs into two-dimensional image data as one-dimensional time series data [10]. By searching for two concave points between touching pigs and connecting the corresponding concave points, the touching pigs could be separated as individual pigs. However, there is a problem where the pigs cannot be separated as individual ones with a conventional separation method when the pigs are overlapped between them. In the case of the overlapping pigs, for example, the shape of the overlapping pigs appears to complex shapes such as X shape and T shape. Because more than two concave points may be obtained in such a complex overlapping shapes, accurate concave

points cannot be obtained which are used to separate the overlapping pigs.

In another studies, a separation method for overlapping people or vehicles in surveillance camera environment has been proposed by using shape information and color information of the objects. In the case of overlapping among people, for example, the overlapping people can be separated as the individuals by using the outer shape of their head [11,12].

In the case of vehicle overlapping problem, some studies have been reported on separating overlapping situations between vehicles using associative tracking method using spatio-temporal information through color information [13]. However, it is difficult to apply this method for the overlapping situations between pigs because this method uses the shape information and the color information of the vehicles and people. For example, the shape information of the head and body of a pig cannot be applied to separate the overlapping pigs because the pig's head and body are not clearly distinguished. Since the color of the pigs is very similar, moreover, the separation method using color information is difficult to be simply exploited for the occlusion situations.

Meanwhile, a separation method using the depth information was proposed by analyzing differences the depth values among the pigs [14-16]. Because of inaccurate depth information, however, the previous methods have a difficulty where the threshold for separating the overlapping pigs should be set continuously according to their posture or position. In this paper, we propose a separation method for overlapping pigs by using a deep-learning technique and some simple image processing techniques.

3. PROPOSED METHOD

In order to separate the overlapping pigs, we perform the foreground detection from the pig room and data augmentation with YOLO in real-time. First, we perform the foreground detection in the pig room by using data fusion of both infrared and depth information before separating the overlapping-pigs. Initially, the region of interest (ROI) in both the input infrared and depth frames is set to only concentrate the activity region of pigs. Then, spatiotemporal interpolation technique is applied to the depth image for removing noises generated from inaccurate depth information. In the next step, the background (e.g., floor and wall) is subtracted from the foreground (i.e., the pigs) by analyzing the depth information, and the contrast of the infrared information is improved with a contrast enhancement technique to roughly localize the pigs. After that, simple image processing techniques are applied to precisely detect the pigs from data fusion (i.e., integrating infrared and depth

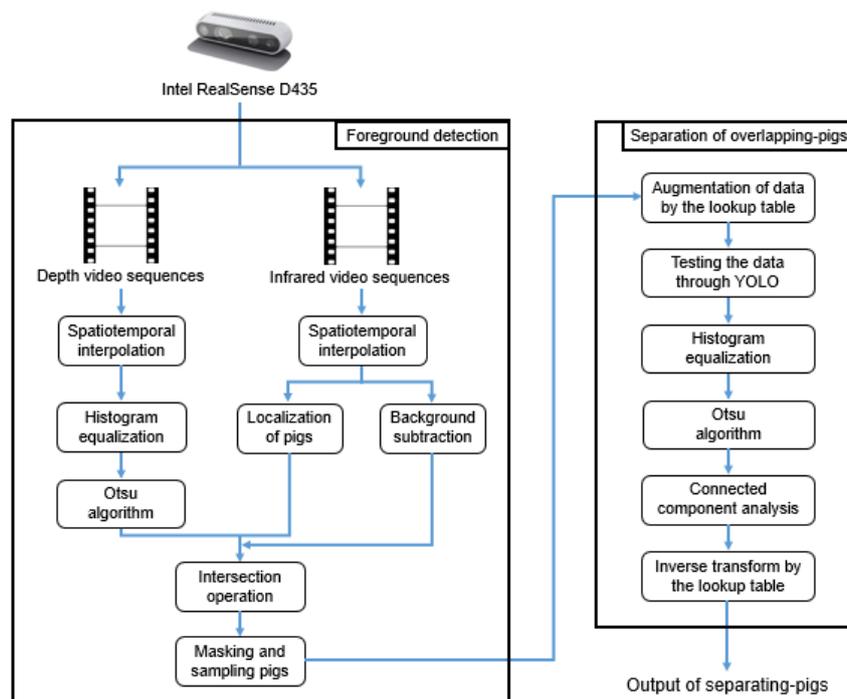


Figure 1. The overall structure of the proposed method.

information). Using the advantages of the data fusion, the pigs can be detected effectively in the low-contrast.

Second, we use YOLO to separate the overlapping-pigs from the result of the foreground detection. Firstly, the input data sampled from the result of the foreground detection is rotated and augmented by the lookup table. In fact, a trigonometric function can be used to rotate the sampled data; however, it is time-consuming to simply utilize the function when rotating the sampled data. By using the pre-computed lookup table for rotation angle, thus, the input data can be rotated and augmented in real-time. Then, YOLO can be used to detect the overlapping-pigs from a bunch of the rotated and augmented data, and some image processing techniques are applied to separate the overlapping-pigs from the YOLO results. Finally, the lookup table can be exploited to inversely transform the result of separated pigs into its original angle. Figure 1 shows the overall structure of the proposed method.

3.1 Preprocessing

First of all, we localize the pigs from the depth information. From an input depth frame, we set the ROI to only focus on the necessary regions from the whole pig room. As preprocessing step, spatiotemporal interpolation [17] with 4×4 window is then applied for removing noises (i.e., undefined pixels). Note that the spatiotemporal interpolation technique should be iteratively performed until the undefined pixels are removed. In order to realize the locations of each pig in the depth frame, the frequencies of the depth values from both background

and foreground are calculated from histogram analysis. We can confirm that the background region is larger than that of the foreground (i.e., the pigs) in the pig room. In other words, the most frequent depth value can be defined as a threshold for background subtraction. Then, the pigs in the pig room can be roughly localized by using the defined threshold. However, some part of the background is not subtracted which has the same depth values with the pigs. For instance, the wall in the pig room may remain through the defined threshold because the height of the wall may be similar to the height of the pig.

In order to remove the remaining wall, the background depth map is modeled from the depth video sequences recorded during 24-h period. First, the floor and other regions (i.e., the wall and pigs) for every frame are respectively divided by using the threshold that is defined by histogram analysis. Here, we define the floor and the other parts as the floor depth map and the other depth map, respectively. Then, the floor depth map and the other depth map are independently updated with the depth values of the floor and the other parts during the 24 h videos. After updating each depth map, the depth values of the other depth map are overlapped to the not updated regions of the floor depth map because the depth values of the floor are only updated from the interpolated depth frame through the threshold.

After the background depth map is completely modeled, frame difference is performed between the background depth map and every interpolated depth frame. Then, the localized pigs can be obtained by using histogram equalization (HE) [18] and Otsu

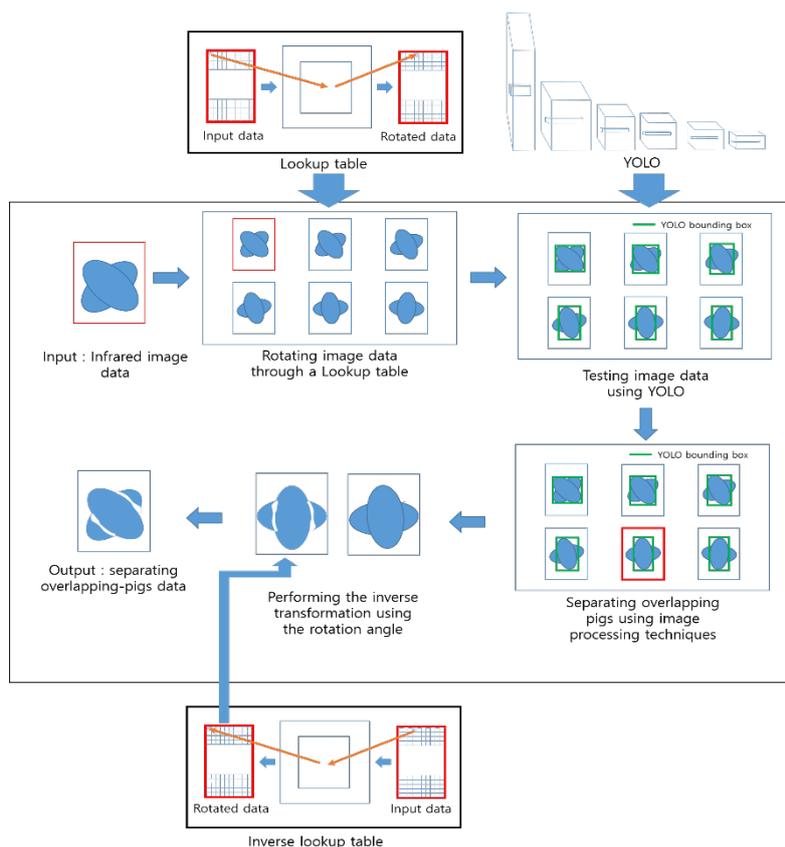


Figure 2. Outline of the separation method.

algorithm [19] on the depth frame. Finally, the depth frame with the localized pigs is used into the infrared frame for robustly detecting the pigs in the low-contrast conditions.

The pigs from infrared information can be detected by analyzing its accurate gray values. Despite of the accurate values of the infrared information, there is a problem that the pigs cannot be detected according to various illumination conditions such as the low-contrast. Because the infrared information may be affected by various illuminations, the depth frame with the localized pigs can be utilized for accurately detecting the pigs regardless of the illumination conditions. First, the ROI of the input infrared frame is also set as the same manner of the depth frame. Then, the histogram equalization (HE) is conducted to solve the low-contrast conditions. In the next step, the Otsu algorithm is applied to roughly localize the pigs after conducting HE into the infrared frame.

Because the contrast of the floor and wall in the infrared frame is also coordinated consistently by applying HE, however, the localized pigs cannot be identified between the floor and wall. Therefore, the pigs localized from each information can be detected with data fusion between the infrared and depth information.

In order to detect the pigs in the pig room, the intersection operation is performed between the infrared and depth frame where the pigs are localized.

Even if the pigs can be detected from both the data, however, the background in the pig room are still detected. For example, the wall and floor are still detected as the foreground because of roughly localized frame with some remaining noises.

In an attempt to only detect the pigs, the frame difference between the input depth frame and the background depth map is used. Only the pigs can be then detected from the intersection operation between the roughly localized frame and the depth frame where the wall and floor are removed through the frame difference. Given the detected frame, the post-processing using some image processing techniques is performed to accurately detect the pigs. In order to remove the remaining noise, an erosion operation is performed to remove and minimize small noises that are adjacent to the objects or generated from the intersection operation. Then, the connected component analysis (CCA) is applied into the detected frame for labeling all of the objects, where the small labeled objects are removed. After removing the noises, the pigs can be detected by using a dilation operation to recover the shapes of the pigs.

3.2 Separation of Overlapping-Pigs

The outline of the separation method for overlapping-pigs consists of four steps as shown in Figure 2: the input data augmentation, the bounding

box selection, the overlapping-pigs separation, and the inverse transform to its original angle.

We use YOLO for separating the overlapping-pigs. YOLO is an object detector that detects objects in real time and separates them with high accuracy. However, since YOLO is a bounding box-based object detector, it has a characteristic to create a bounding box parallel to the x-axis. This characteristic causes some problems when separating the overlapping-pigs because of the shapes of the various occlusion. As shown in Figure 3, for example, a bounding box is generated on the overlapping-pigs through YOLO for X shape and T shape. In this type of occlusion, we tried to detect only occluding pig, but the area of the interest of the bounding box also include occluded pig. To solve this problem, we augment the data at the test-time. There are methods such as shifting, flipping, and rotation to augment data. Among such the augmentation categories, we use data rotation from the categories to overcome the characteristic of the bounding box.

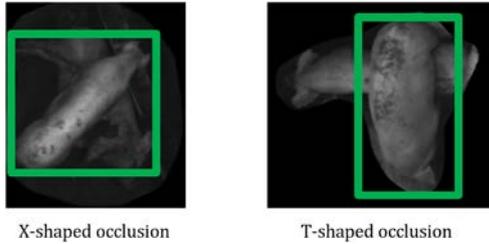


Figure 3. Various bounding boxes obtained from YOLO in case of occlusion.

The image can be rotated through the trigonometric function. In this case, however, it is hard to satisfy the real time because of a relatively large amount of computing loads. In order to satisfy the real-time processing, we rotate the data using a pre-defined lookup table. The lookup table rotates the data clockwise. The rotation angle is from 0° to 50° , and it generates six data including the input data (0°). The six images are combined into one image in order to perform YOLO efficiently.

In the next step, the optimally-rotated YOLO result is obtained according to the area of the margin. Here, a “margin” is defined as the area excluding the occluding pig in the bounding box. The area of the margin can be calculated using the following equation (1):

$$margin = \frac{P_{total} - P_{occluding_pig}}{P_{total}} \quad (1)$$

where P_{total} means the number of all pixels in the bounding box, and $P_{occluding_pig}$ means the number of occluding pig pixels. Margins are calculated on all the augmented data to obtain the area of the margin for each of the six data including the input data. The area of the six margins obtained are compared and the

bounding box of the largest area of margin is selected as the optimally-rotated YOLO result. That is, the data, which is the area of the largest margin, can be determined as the representative data among the six augment data.

Then, we apply some image processing techniques to the optimally-rotated YOLO result in order to separate the overlapping-pigs. First, histogram equalization (HE) technique [18] is applied to the optimally-rotated YOLO result in order to clearly distinguish the boundary line between the overlapping pigs. When the histogram smoothing is applied, the intensity distribution of the whole image using the smoothing becomes uniform, and thus the contrast between the two pigs can be improved. As a next step, Otsu technique [19] is applied in order to separate the two pigs. Then, the binary result from Otsu is labeled with connected component analysis technique [20], and the number of foreground pixels is calculated to derive the size of each connected component. By selecting the largest and the second largest connected component, we can obtain the occluding pig and the occluded pig as a final result of image processing techniques (i.e., some image noise can be removed).

Finally, we use an inverse lookup table to rotate the image processing result back to its original angle. That is, the inverse lookup table is a pre-computed table that rotates the pixel locations of the image processing result in the opposite direction with the lookup table. As a result, we obtain the final separation result of the overlapping-pigs. Algorithm 1 describes the overall proposed method.

Algorithm 1 Separation of overlapping-pigs

Input: Infrared and depth information sequences

$I_{mask} = \text{FG_Detect}(I_{infrared}, I_{depth}, I_{bg});$
 $I_{sample} = \text{SamplingPigs}(I_{mask}, I_{infrared});$

for $i = 1$ to 5 :

$I_{rotate}^i = \text{lookup}_i(I_{sample});$

$I_{optimal} = \text{CompareMargines}(I_{rotate});$

$I_{bbox} = \text{YOLO}(I_{optimal});$

Apply histogram equalization into I_{bbox} ;

Separate overlapping pigs in I_{bbox} with Otsu;

Post-process I_{bbox} through CCA;

$I_{sep} = \text{lookup}^{inverse}(I_{bbox});$

Output: Separating overlapping pigs into individual pigs

4. EXPERIMENTAL RESULTS

The following experimental setup was used to conduct the proposed method: Intel® Core™ i7-7700K, NVIDIA GeForce GTX1080 Ti, 32GB RAM, and OpenCV 3.4. For the experiment, an Intel RealSense D435 camera was installed on the ceiling of 3.2m above the bottom. There were 9 pigs in the room, and we obtained video frame images having resolution size of 1280×720 pixels. In the case of YOLO, we produced a model through the training data, which was composed of 774 frames. The YOLO learning parameters were 0.001 for learning rate, 0.9 for momentum, and 0.0005 for decay. The activation function was leaky ReLU function. We then used 47 test frames that were not used as training data.

We used the depth and infrared video sequences obtained from Intel RealSense camera during a 24 h period. As explained in Section 2, various illumination conditions in the infrared and depth videos were confirmed such as low contrast. In particular, the low-contrast conditions were evidently identified when the pigs were located at the corners in the pen. Thus, we detected the pigs while considering the conditions.

Initially, we modeled the background depth map as an independent procedure for conducting the frame difference into the input depth frame. Then, the spatiotemporal interpolation technique was applied to the 1296 frames extracted from each video. Note that because the spatiotemporal interpolation technique was interpolated from three frames to one frame, 1296 frames were needed to detect the pigs in 432 frames. With each interpolated frame, simple image processing techniques were conducted to each domain.

In the case of depth information, a histogram analysis was performed for background subtraction in the input depth frame. The frequency of the depth value corresponding to the background converged to 53, and the depth threshold for segmenting the background was defined as 53. The background-subtracted depth frame was then derived using the threshold defined through histogram analysis. In addition, the frame difference between the background depth map and the interpolated depth frame was carried out. In the case of the procedure of infrared information, the Otsu algorithm was used to define the parameter for segmenting the background for roughly localizing the pigs. Using the background-subtracted depth frame and each localized frame from each information, the pigs could be detected by performing the intersection operations among the frame attributes.

For the detected frame, the morphology operation and CCA were conducted as the post-processing steps for refining the detected pigs. As the size of each noise calculated by CCA was less than 100, the noises were simply removed with the threshold defined as 100. After that, a dilation operation was conducted three

times to sufficiently recover the shape of the pigs, and as a result, all of the pigs in the pen could be accurately detected.

Then, we sampled the input data from the results of the foreground detection, which was used to rotate counterclockwise for selecting an optimal bounding box. The optimal bounding box was selected when the occluding pig was aligned with x-axis or y-axis. Accordingly, the rotation angle of the test-time augmentation was set to 0° ~ 50° with 10° interval. Then, the six rotated data were merged into a single synthetic image. Through this synthesis, six YOLO executions could be reduced to one YOLO execution (i.e., YOLO is You Only Look Once, independent of the number of objects to be detected). In addition, we used a pre-computed lookup table, instead of applying a trigonometric function directly, to reduce the execution time. In the case of rotating one data using the trigonometric function, the execution time of 27 msec was required. On the other hand, the lookup table could reduce the rotation time to less than 1msec.

Fig. 4 shows the YOLO results after rotating 0°, 10°, 20°, 30°, 40°, and 50° for X-shaped and T-shaped overlapping pigs, respectively.

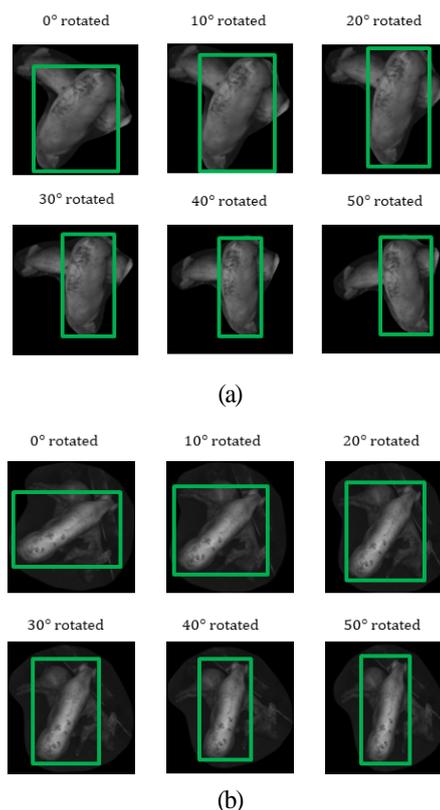


Figure 4. Results of YOLO from the rotated data. (a) Results of YOLO from the rotated T-shaped data and (b) Results of YOLO from the rotated X-shaped data.

Then, we measured the horizontal and vertical lengths of the six bounding boxes, and obtained the horizontal and vertical ratios of the bounding boxes. The YOLO result with the largest ratio was selected as the optimally-rotated YOLO result. In Figure 4(b), for example, the ratios of each rotation (0° rotation) were 1.05 (for 0° rotation), 1.29 (for 10° rotation), 1.62 (for 20° rotation), 1.80 (for 30° rotation), 2.05 (for 40° rotation), and 2.27 (for 50° rotation), respectively. Then, the largest ratio was 2.27, and thus the YOLO result rotated by 50° rotation was selected as the optimally-rotated YOLO result.

With the optimally-rotated YOLO result, we applied some image processing techniques to separate the overlapping pigs. Histogram smoothing was first applied to the optimally-rotated YOLO result in order to clarify the intensity value of the overlapping-pigs by improving the contrast of the optimally-rotated YOLO result. Then, Otsu was applied to separate the individual pigs from the overlapping-pigs. After the connected component analysis with the Otsu's result, we computed the size of each connected component. Finally, we selected two largest connected components as the occluding pig and the occluded pig, and rotated the result to the original angle by using the inverse lookup table.

In order to evaluate the separation result of the proposed method qualitatively, we compared it with the state-of-the-art deep learning-based instance segmenter (i.e., Mask R-CNN [21]) and YOLO [6] only (i.e., applying YOLO to overlapping-pigs directly without any rotation). As shown in Figure 5, Mask R-CNN could not distinguish the overlapping-pigs depending on the overlapping patterns, whereas YOLO only could not distinguish the overlapping-pigs at all with the selected frames. On the contrary, the proposed method could distinguish the overlapping-pigs with the selected frames.

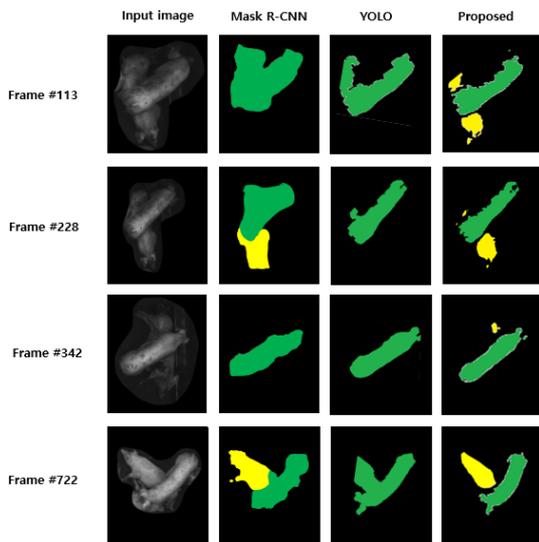


Figure 5. Separation results of overlapping-pigs.

In order to evaluate the separation performance of overlapping-pigs quantitatively, the results of the three methods were compared. As a performance metric for accuracy, we compared the separation results with ground-truth at pixel-level. Each accuracy of the methods was calculated by equation (2):

$$Accuracy = \frac{N_{TP}}{N_{TP} + N_{FP} + N_{FN}} \quad (2)$$

where true positive (N_{TP}) is defined to a pixel on the separated pigs, false positive (N_{FP}) is defined to a pixel on the background as the separated pigs, and false negative (N_{FN}) is defined to a pixel on the separated pigs as the background.

Since real-time performance is important in any surveillance system, we compared the execution times. Finally, these two performance metrics (i.e., accuracy and time) are generally conflicting, we computed a “collective” performance metric defined as accuracy/time.

Table 1 shows the comparison results of Mask R-CNN, YOLO only, and the proposed method. The proposed method could provide better accuracy than Mask R-CNN and much better accuracy than YOLO. Also, the proposed method could provide much faster execution time than Mask R-CNN, while it could provide slightly slower execution time than YOLO due to the additional image processing steps. With the accuracy/time performance metric, the proposed method could improve the collective performance of Mask R-CNN by a factor of 11, and that of YOLO by 27%.

Even though the overlapping pigs were separated through the proposed method in real-time, we should consider the occlusion issue for separating among more than two pigs. The occlusion among the numerous pigs in a complex pig room should be solved with the extended separation method in order to manage the pigs' health care. Therefore, we will extend the proposed method to precisely separate occlusion among more than two pigs.

Table 1. Performance comparison

Methods	Accuracy (%)	Time (msec)	Accuracy / Time (%/msec)
Mask R-CNN[21]	79.87	254.35	0.31
YOLO[6]	55.61	20.42	2.72
Proposed	83.33	24.01	3.47

5. CONCLUSION

Separation of dense pigs in a crowded environment is an important issue to automatically manage pig

farms. Recently, many studies have been reported to detect pigs with YOLO (i.e., one of the fast deep learning-based object detectors). With the axis aligned bounding box-based method, however, separating overlapping-pigs is difficult, depending on the complicated occlusion patterns.

In this study, we proposed the real-time separation method for overlapping-pigs using test-time augmentation with the effective foreground detection in the pig room. First of all, we applied the spatiotemporal interpolation into the depth information for removing noises and roughly localizing the pigs in the pig room. Then, the infrared information was concurrently utilized to the depth information as data fusion, and we simply applied effective image processing techniques to precisely detect the pigs in real-time.

Through the results of the foreground detection, we used YOLO and image processing techniques by using the pre-computed lookup table for test-time augmentation. As a result of these procedures, the overlapping-pigs could be separated fast and accurately regardless of the complicated occlusion patterns.

Experimental results show that the overlapping-pigs could be separated with accuracy of 83.33% and the execution speed of 24.01msec. These results show that the accuracy/time performance of the proposed method was 1,119% higher than that of Mask R-CNN, and it was 27% higher than that of YOLO only. Future studies will be carried out on a parallel-processing method that can handle the whole process of individual pig analysis in real-time.

6. ACKNOWLEDGMENTS

This research was supported by the Basic Science Research Program through the NRF funded by the MEST (2018R1D1A1A09081924) and the Leading Human Resource Training Program of Regional Neo Industry through the NRF funded by the MSIP (2016H1D5A1910730).

7. REFERENCES

- [1] Y. Chung, S. Oh, J. Lee, D. Park, H. Chang, and S. Kim, "Automatic Detection and Recognition of Pig Wasting Diseases Using Sound Data in Audio Surveillance Systems," *Sensors*, Vol. 13, No. 10, pp. 12929-12942, 2013.
- [2] A. Wongsriworaphon, B. Arnonkijpanich, and S. Pathumnakul, "An Approach Based on Digital Image Analysis to Estimate the Live Weights of Pigs in Farm Environments," *Computers and Electronics in Agriculture*, Vol. 115, No. C, pp. 26-33, 2015.
- [3] M. Oczak, K. Maschat, D. Berckmans, E. Vranken, and J. Baumgartner, "Automatic Estimation of Number of Piglets in a Pen During Farrowing, Using Image Analysis," *Biosystems Engineering*, Vol. 151, pp. 81-89, 2016.
- [4] M. Ju, H. Baek, J. Sa, H. Kim, Y. Chung, and D. Park, "Real-Time Pig Segmentation for Individual Pig Monitoring in a Weaning Pig Room," *Journal of Korea Multimedia Society*, Vol. 19, No. 2, pp. 215-223, 2016.
- [5] L. Lee, L. Jin, D. Park, and Y. Chung, "Automatic Recognition of Aggressive Behavior in Pigs Using a Kinect Depth Sensor," *Sensors*, Vol. 16, No. 5, pp. 631, 2016.
- [6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real Time Object Detection," *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779-788, 2016.
- [7] J. Seo, M. Ju, Y. Choi, J. Lee, Y. Chung, and D. Park, "Separation of Touching Pigs Using YOLO-Based Bounding Box," *Journal of Korea Multimedia Society*, Vol. 21, No. 2, pp. 77-86, 2018.
- [8] M. Ju, Y. Choi, J. Seo, J. Sa, S. Lee, Y. Chung, and D. Park, "A Kinect-based Segmentation of Touching-Pigs for Real-Time Monitoring," *Sensors*, Vol. 18, No. 6, pp. 1746, 2018.
- [9] Intel RealSense D435, Intel. Available online: <https://click.intel.com/intelr-realsensetm-depth-camera-d435.html> (accessed on 28 Feb. 2018).
- [10] H. Baek, Y. Chung, M. Ju, Y. Chung, D. Park, and H. Kim, "Pig Segmentation using Concave Points and Edge Information," *Journal of Korea Multimedia Society*, Vol. 19, No. 8, pp. 1361-1370, 2016.
- [11] Y. Do, "Dividing Occluded Humans Based on an Artificial Neural Network for the Vision of a Surveillance Robot," *Journal of Institute of Control, Robotics and Systems*, Vol. 15, No. 5, pp. 505-510, 2009.
- [12] H. Choi, S. Hong, and J. Ko, "Merge and Split of Players under MeanShift Tracking in Baseball Videos," *Journal of Advanced Navigation Technology*, Vol. 21, No. 1, pp. 119-125, 2017.
- [13] J. Lim, S. Kim, C. Lee, and M. Lee, "Overlap Removal and Background Updating for Associative Tracking of Multiple Vehicle," *Journal of KIISE*, Vol. 16, No. 1 pp. 90-94, 2010.
- [14] H. Lee, Y. Choi, J. Sa, Y. Chung, and D. Park, "Detection of Occluding Pigs Using Depth Information in a Pigsty," *Proceeding of the Fall Conference of the Korea Multimedia Society*, Vol. 25, No. 2, pp. 833-835, 2018.

- [15] H. Lee, H. Lee, J. Kim, Y. Choi, H. Kim, Y. Chung, D. Park, and H. Kim, "Occluded-Pigs Detection and Separation Using Depth Information," *Proceeding of the Fall Conference of the Korea Multimedia Society*, Vol. 20, No. 2, pp. 813-815, 2017.
- [16] H. Seo, H. Lee, C. Park, J. Seo, Y. Chung, D. Park, and H. Kim, "Occluding Pigs Individual Detection Using Depth Information," *Proceeding of the Conference of the Workshop on Image Processing and Image Understanding*, 2018.
- [17] J. Kim, Y. Chung, Y. Choi, J. Sa, H. Kim, Y. Chung, D. Park, and H. Kim, "Depth-based Detection of Standing-Pigs in Moving Noise Environments," *Sensors*, Vol. 17, No. 12, pp. 2757, 2017.
- [18] M. Eramian and D. Mould, "Histogram Equalization using Neighborhood Metrics," In *Proc. of the 2nd Canadian Conference on Computer and Robot Vision (CRV'05)*, pp. 397-404, 2005.
- [19] N. Otsu, "A Threshold Selection Method from Gray-level Histograms," *IEEE Trans. Syst. Man Cybern*, Vol. 9, No. 1, pp. 62-66, 1979.
- [20] K. Pulli, A. Baksheev, K. Korniyakov, and V. Eruhimov, "Real-time Computer Vision with OpenCV," *Communications of the ACM*, Vol. 55, No. 6, pp. 61-69, 2012.
- [21] K. He, G. Gkoxari, P. Dollár, and R. Girshick, "Mask R-CNN," *Proceeding of the IEEE International Conference on Computer Vision*, pp. 2980-2988, 2017.