Health Monitoring of Group-Housed Pigs using Depth-Enabled Multi-Object Tracking

Mateusz Mittek, Eric T. Psota, Lance C. Pérez
Department of Electrical and Computer Engineering
University of Nebraska-Lincoln
Lincoln, NE 68588, USA
{mmittek, epsota, lperez}@unl.edu

Ty Schmidt, Benny Mote
Department of Animal Science
University of Nebraska-Lincoln
Lincoln, NE 68588, USA
{tschmidt4, tmote8}@unl.edu

***Abstract*—A practical, low-cost system is presented for continuous tracking of animals using depth-enabled multi-object tracking. The system is capable of producing detailed, long-term, continuous 3D movement data that can be used to detect eating/drinking, aggression, and a multitude of other social interactions. Results also demonstrate that physical parameters like weight can be reliably estimated by the system. The combination of movement data and physical parameters make it possible for the system to monitor growth rate, classify aggression, and detect early signs of compromised health allowing for individualized care and management in large group settings.**

I. Introduction

One of the biggest challenges to ensuring the wellbeing and efficiency of pigs is rapidly and accurately identifying compromised (sick or injured) pigs. To date, the only method available for identification of compromised pigs is via manual observation for visible indicators of sickness or illness (clinical symptoms). However, given the quantity of pigs in modern group-housed settings, it is a daunting task to ensure that each pig is visually inspected even as frequently as once a day.

The pork industry is currently operating with a piglet loss rate of 20.8% from birth to weaning when stillborns are included [9]. According to Metafarms data analysis, first quarter 2016 is seeing a nursery mortality rate of about 4% at early stages of animal development [11]. In addition, the industry is dealing with the devastating impact of multitude of other viruses and bacteria such as PRRS, PEDv and PCV2. PRRS alone is estimated to result in an annual loss of $664 million to U.S. swine producers in part due the to loss of production efficiency [7]. PRRS infection slows growth by 20% and harms feed conversion by 15%, primarily in the early phases of the infection [4]. Altogether, this data indicates that early identification and treatment can yield substantial benefits to producers.

A depth-enabled vision system is presented here as an alternative to manual observation. The system automatically identifies, maintains identification, continuously tracks movement, and estimates physical properties of pigs within a commercial setting. Figure 1 illustrates the individualized 3D tracking of each pig's location, orientation, and activity data. While the system was designed for swine, it can be applied to other animals, including livestock, companion animals, captive wildlife, and biomedical animals.

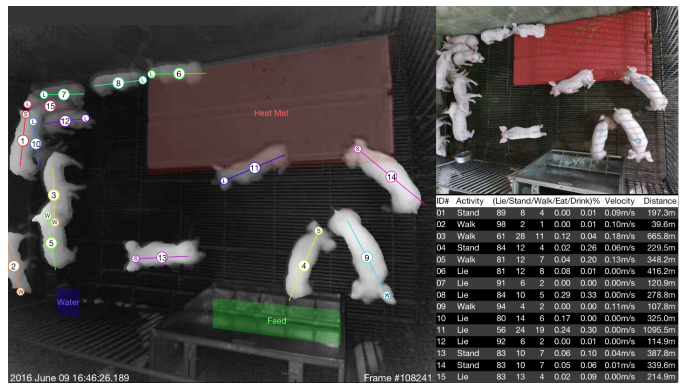

Figure 1. Tracking software vizualization near the end of the trial; displays current activities, velocity, locations of each pig, total travelled distance and the cumulative time that each pig has spent performing each activity.

II. Background

Over the past two decades, researchers have studied a variety of methods and technologies used for behavioral monitoring of animals [2], [15]. A popular approach to behavioral monitoring is to equip the animals with wearable activity-tracking devices. For example, accelerometer data-logging collars have been used to identify the activities of pigs [3] and to measure the level of locomotion in sows as they approach farrowing [2]. Using machine learning, these systems are able to classify feeding, rooting, walking, lying laterally, and lying sternally with an average accuracy of 75%.

While accelerometer data-logging devices are capable of detecting types of locomotion and behavior, there are several real-world drawbacks. Accelerometers must be physically removed from the pig for data collection prior to processing, introducing additional stress to the animals. Sensors can also be lost or destroyed due to biting.

Automated video tracking presents a non-invasive alternative to wearable devices. By using traditional color imaging and background subtraction, researchers have designed methods for tracking in constrained environments where pigs walk individually in front of the camera [10]. Cameras have also been used to monitor locomotion and behavior on individual animals within a narrow camera view [1].

To monitor multiple pigs simultaneously, it is necessary to segment them from both the background and from one

Figure 2. Color, infrared, and depth frames captured by the Kinect v2 camera.

another; a difficult task considering their tendency to lie or stand in groups (see Figure 1 (top-left corner)). Previous attempts to achieve this level of segmentation using supervised learning produce accurate results when the pigs are all simultaneously viewable (i.e., non-occluded) from a top-down camera [13]. Recently, researchers developed a system to automatically monitor locomotion by tracking multiple pigs in the same environment by fitting predefined shapes resembling a pig's silhouette from a vertical perspective to the regions obtained through conventional background subtraction [8]. These shapes allow the system to separating abutting pigs and identify their orientation. Their system is able to achieve 90% accuracy when classifying pigs as active or inactive.

Computer vision has also been used to identify the lying behavior of group-housed pigs as a function of temperature [12] and the movement patterns of individual pigs and the entire herd have been extracted through optical flow to detect abnormal behaviors [6]. The authors do not attempt to perform long-term tracking, and acknowledge the high computational complexity of trying to track individual animals.

Recognizing the need for a tracker that can handle multiple homogenous targets, Giancardo et al. attempt to achieve long term tracking of multiple mice in a confined living space through the processing of thermal images [5]. They define a set of social behaviors that correlate with certain genetic conditions, introduce image processing techniques to distinguish between abutting mice, and achieve individualized tracking and behavioral monitoring. Their system, however, still swaps labels between mice approximately once every 30 seconds.

A key disadvantage when using conventional images for the initial stage of background subtraction is that segmentation failure occurs when foreground colors resemble background colors. Changes in the image resulting from shadows and fluctuations in lighting often result in mislabeled foreground objects [16]. In contrast, when using a depth sensor, the foreground can often easily be identified as areas that are closer to the camera than the background. Thus, without altering the background or camera pose, foreground segmentation of depth frames is inherently more robust.

In 2013, a group of researcher at the Princeton Vision Group recognized the importance of depth information and established an online benchmark to evaluate the tracking performance using depth-enabled cameras [14]. However, the benchmark does not consider the challenge of tracking multiple visually indistinguishable targets over long time periods, and most methods listed on the benchmark rely at least partially

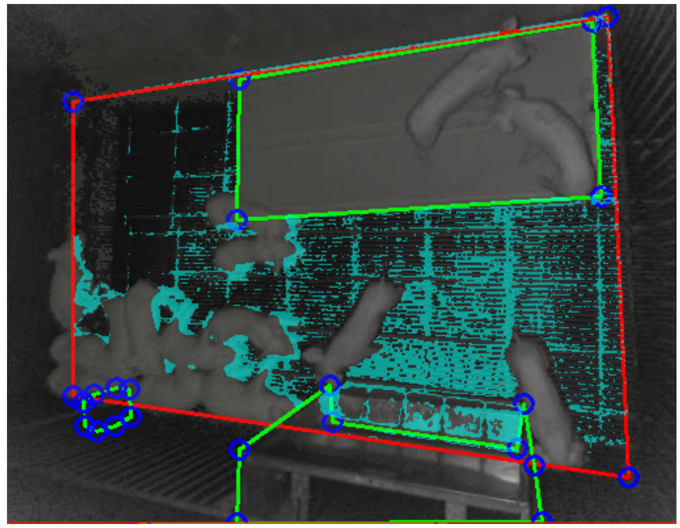

Figure 3. User-selected regions corresponding to the pen floor (red boundary) and the heat mat, feeder, and waterer (green boundaries), plane estimation using RANSAC (cyan). The cyan points correspond to the dominant plane that lies within the pen floor region defined by the user.

on the color distributions of targets. Thus, current solutions are not sufficient for robust long-term animal tracking. As an alternative to conventional multi-object tracking, the method presented here uses 3D shape fitting and motion modeling to track multiple homogenous targets over long durations

III. Method

Modern consumer-grade real-time depth cameras provide a viable solution to serve as the backbone of an enhanced visual tracking system [14]. A popular choice of such camera is the Kinect v2 gaming peripheral developed by Microsoft to track human movement. The Kinect v2 comes equipped with a high-definition camera, an infrared illuminator, and a time-of-flight depth sensor that produces color, infrared, and depth frames, as illustrated in Figure 2. In addition to facilitating depth measurement, the infrared illuminator makes it capable of tracking day and night without the need for visible light.

The proposed method processes the frames provided by the Kinect v2 camera in order to track multiple pigs simultaneously in a group-housed environment. The stages of processing required for movement and activity tracking are described in the following sections.

A. Adaptive capture of Kinect frames

Often, it is a waste of computational resources to capture new data during times of inactivity, as young pigs spend the majority of their time sleeping. Therefore, the system is designed to track total movement between the last recorded depth frame and the current depth frame provided by camera. Specifically, when the percentage of pixels in the depth frame that change by 5 cm or more between the previously captured depth frame and the current one exceeds 0.2%, new data is capture. For static scenes without reflective surfaces, observed errors in the Kinect v2 depth measurements rarely exceed 3cm. Thus, measurements that change by 5cm or more between subsequent frames nearly always indicate true movement within the scene. This value was empirically determined to provide a good balance between smooth motion and a reduced computational load.

A one-time visual calibration of the system requires the user to select corner points defining the pen boundaries, feeder, waterer, and heat mat (see Figure 3). Accurate identification of the pen floor bounds is critical for foreground isolation and tracking. To correct manual selection errors, a random sample and consensus (RANSAC) plane-fitting routine is applied to all depth points that lie within the bounds of the area selected by the user (minus those belonging to the heat mat, since it is 1" higher than the pen floor).

By taking into account internal camera parameters and the true dimensions of the pen floor, the four corners enable the 6-degree-of-freedom pose of the camera to be established with respect to this plane, thus enabling the software to reorient the 3D points detected by the Kinect v2 camera from the camera's local coordinate system to the coordinate system of the pen floor, where the bottom left corner of the pen is $(x, y, z) = (0, 0, 0)$ and the z-axis is perpendicular to the pen floor.

B. Back point extraction

Although the Kinect v2 captures a large number of depth measurements corresponding to each pig, for the purpose of robust tracking it was necessary to identify an area of the pig that is nearly always visible and stable with respect to the pigs' location and orientation. With the camera mounted vertically above the pen, the points along the back were observed to satisfy both of these conditions.

To detect points lying on the back, the normal direction of each point in the foreground is calculated. Points with the Z components of the unit normal vector greater than 0.75 belong to surfaces roughly perpendicular to the ground plane and identified as back points. Points must also satisfy convexity criteria (as viewed from the camera), since the body shape of pigs conforms roughly to an ellipsoid.

C. Appearance modeling using ellipsoid fitting and state-conditioned motion filtering

Once the 3D point clouds have been processed so that they include only the points along the back, an ellipsoid tracker is used to maintain the position and orientation of the pigs. The tracker is initialized with the pig locations manually and, in all future frames, adjusts the position of an ellipsoid to each new set of back points. It is essentially a variant of mean shift, where the kernel is an ellipsoid, and points are associated with the means in order from most reliable to least reliable.

Finally, a state-conditioned Kalman filter is used to estimate the position and orientation of each pig. The Kalman filter has two modes of operation, conditioned on whether the pig's previous height indicates that it is lying down or standing up. When the pig is lying down, horizontal positioning relies heavily on the previous state, discouraging motion. When the pig is standing, horizontal positioning relies more on the observed motion via ellipsoid fitting. In practice, this two-state approach addresses the problem of object occlusion and

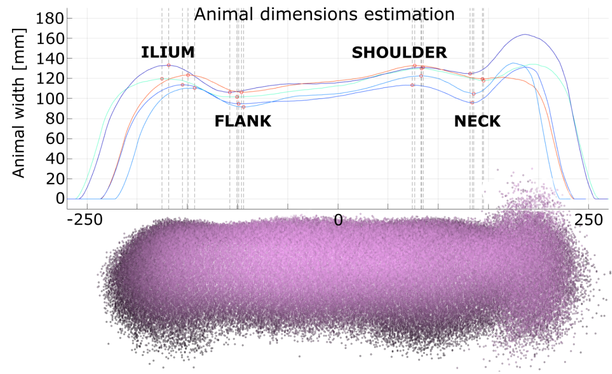

Figure 4. Point cloud accumulation for physical dimension estimation (bottom) along with width estimates of ilium, flank, shoulder, and neck (top).

allows pigs to cross over each other while maintaining their identification.

D. Weight estimation from point clouds

Recognizing that equipping each pig with unique ear tags and/or paint patterns places an unwanted burden on a commercial facility, a more desirable solution would be to differentiate each pig based on its physical characteristics. In an effort to extract the physical dimensions of the pigs, point cloud data collected by the system over long durations is accumulated as shown in Figure 4. Point cloud accumulation is performed by: 1) identifying when the pig is standing, 2) transforming the points to a common coordinate system by undoing shift and rotation, and 3) cleaning up noise by thresholding in a voxel grid. The model extracted through point cloud accumulation produces the top-down side contours illustrated in Figure 4.

IV. Results

Preliminary results indicate that the proposed tracking method is capable of reliably tracking the movements of multiple pigs simultaneously over long durations of time. In an experiment designed to evaluate the system's capabilities, the Kinect v2 camera connected to an Intel NUC computer was mounted above a single pen at Union Farms in Ulysses, NE. Fifteen pigs were then moved into the pen and the computer captured 118,411 frames at an average of 6.27 frames per second (fps) for approximately 5 hours and 14 minutes. Note that, using the method described in Section III-A, the frame rate varied between 0 to 20 fps.

The amount of individual and social activity peaks when pigs first enter a new environment with new pen mates. These movement patterns and complex interactions between pigs caused the tracking to produce an error, on average, every 5 minutes. Therefore, to achieve reliable tracking for the full 5 hours and 14 minutes, a video scoring interface was developed to allow a human observer to pause tracking and correct mistakes. Our preliminary experiment required the user to make a total of 72 adjustments the tracking, where they were able to validate the tracking by cross referencing with numeric labels (1-15) painted on pigs' backs.

A sample frame from the tracking visualization is shown in Figure 1. Distinctly colored lines denote the location and

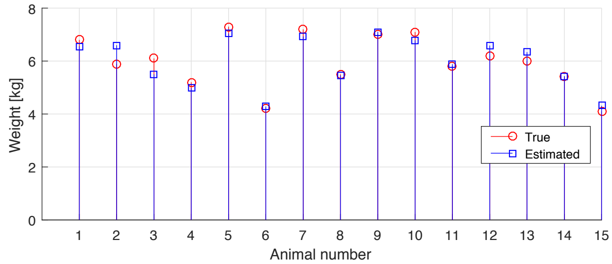

Figure 5. Weight estimation based on visual clues results width mean absolute error of 0.26kg, p-value $= 1.55 \cdot 10^{-6}$, and $r = 0.89$.

orientation of each pig, where the large circle at the center contains the pig ID and the small circle at the head contains an abbreviation of the current activity. The frame number and time stamp are shown on the bottom and the activity distribution and total movement statistics are given in the table on the bottom right quadrant of the images. While there were no health outcomes associated with the preliminary trial, the system tracks activities that are clearly related to health including eating and drinking durations, time elapsed before first meal, and the ratio of activity to inactivity.

To evaluate the system's ability to accurately parameterize pigs, the manually recorded weight of each pig and the weight predicted from the data are presented in Figure 5. An equation derived using least squares relating the measurement extracted from the point cloud data to the true weight is given by

$$\text{Weight} = 0.0002 \times \text{FlankWidth}^2 + 0.04 \times \text{NeckToIlium} - 8.07,$$

where lengths are in millimeters and weight is in kilograms. The mean absolute error between the manually measured weight and the predicted weight is 0.26kg.

Automatically extracted parameters like weight can be used to resolve label swaps and other errors that occur during tracking. They can also be useful health status indicators, such as when there is a loss of weight due to lethargic behavior associated with morbidity. More tracking data will likely reveal additional measurements and statistics that allow the system to uniquely identify each pig in order to recover from label swaps. If successful, this could remove the need for fiducial markers and, as a result, remove the additional stress placed on the animal upon entry to the commercial facility.

V. Conclusion

The proposed visual tracking method was introduced and evaluated in an industrial swine facility. Tracking of multiple group-housed pigs for long durations of time is challenging due to the near-identical appearance of the pigs and their complex movements and interactions. Multiple stages of depth frame processing were presented for extracting the data needed for continuous tracking, and the results demonstrate that the system is capable of achieving continuous tracking of multiple group-housed pigs for long durations of time.

The system invariably loses track of targets and swaps identities due to factors such as dynamic movements and interactions among pigs, momentary system failure, and humans entering the pen. Therefore, a potential solution to the swapping problem is presented that uses automated characterization of physical properties. With the application of additional parameter extraction, it may be possible for the system to automatically correct swaps and allow it to achieve continuous tracking for several weeks. This longitudinal data could eventually be used to predict health outcomes and drastically improve the efficiency of industry practices.

Acknowledgment

The authors would like to thank Lukas Fricke and Union Farms in Ulysses, Nebraska for their kind cooperation and the use of their facilities.

References

[1] A. Cangar, T. Leroy, M. Guarino, E. Vranken, R. Fallon, J. Lenehan, J. Mee, and D. Berckmans. Automatic real-time monitoring of locomotion and posture behaviour of pregnant cows prior to calving using online image analysis. *Computers and electronics in agriculture*, 64(1):53–60, 2008. 1

[2] C. Cornou, J. Vinther, and A. R. Kristensen. Automatic detection of oestrus and health disorders using data from electronic sow feeders. *Livestock Science*, 118(3):262–271, 2008. 1

[3] H. J. Escalante, S. V. Rodriguez, J. Cordero, A. R. Kristensen, and C. Cornou. Sow-activity classification from acceleration patterns: a machine learning approach. *Computers and electronics in agriculture*, 93:17–26, 2013. 1

[4] N. K. Gabler, W. P. Schweer, J. F. Patience, L. Karriker, C. Sparks, G. Gourley, M. Fitzsimmons, K. Schwartz, and T. E. Burkey. The impact of PRRSV on feed efficiency, digestibility and tissue accretion in grow-finisher pigs. *Allen D. Leman Swine Conference 40*, pages 135–136, 2013. 1

[5] L. Giancardo, D. Sona, H. Huang, S. Sannino, F. Manago, D. Scheggia, F. Papaleo, and V. Murino. Automatic visual tracking and social behaviour analysis with multiple mice. *PloS one*, 8(9):e74557, 2013. 2

[6] R. Gronskyte, L. H. Clemmensen, M. S. Hviid, and M. Kulahci. Monitoring pig movement at the slaughterhouse using optical flow and modified angular histograms. *Biosystems Engineering*, 141:19–30, 2016. 2

[7] D. J. Holtkamp, J. B. Kliebenstein, E. Neumann, J. J. Zimmerman, H. Rotto, T. K. Yoder, C. Wang, P. Yeske, C. L. Mowrer, and C. A. Haley. Assessment of the economic impact of porcine reproductive and respiratory syndrome virus on United States pork producers. *Journal of Swine Health and Production*, 21(2):72, 2013. 1

[8] M. A. Kashiha, C. Bahr, S. Ott, C. P. Moons, T. A. Niewold, F. Tuyttens, and D. Berckmans. Automatic monitoring of pig locomotion using image analysis. *Livestock Science*, 159:141–148, 2014. 2

[9] R. Ketchem and M. Rix. What does 780,000 services on high producing farms say?, 2016. 1

[10] N. M. Lind, M. Vinther, R. P. Hemmingsen, and A. K. Hansen. Validation of a digital video tracking system for recording pig locomotor behaviour. *Journal of neuroscience methods*, 143(2):123–132, 2005. 1

[11] Metafarms Analytics Team. Q1 nursery groups see higher mortality, 2016. 1

[12] A. Nasirahmadi, U. Richter, O. Hensel, S. Edwards, and B. Sturm. Using machine vision for investigation of changes in pig group lying patterns. *Computers and Electronics in Agriculture*, 119:184–190, 2015. 2

[13] M. Nilsson, H. Ardö, K. Åström, A. Herlin, C. Bergsten, and O. Guzhva. Learning based image segmentation of pigs in a pen. In *Visual observation and analysis of Vertebrate And Insect Behavior 2014*, pages 1–4, 2014. 2

[14] S. Song and J. Xiao. Tracking revisited using rgbd camera: Unified benchmark and baselines. In *Proceedings of the IEEE international conference on computer vision*, pages 233–240, 2013. 2

[15] A. Stukenborg, I. Traulsen, B. Puppe, U. Presuhn, and J. Krieter. Agonistic behaviour after mixing in pigs under commercial farm conditions. *Applied Animal Behaviour Science*, 129(1):28–35, 2011. 1

[16] Y. Wu, J. Lim, and M.-H. Yang. Online object tracking: A benchmark. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2411–2418, 2013. 2