Process Mining on Video Data*

Arvid Lepsien[1], Jan Bosselmann[1], Andreas Melfsen[2], and Agnes Koschmider[1]

[1] Group Process Analytics, Computer Science Department
Kiel University, Germany
{stu201368|stu206958}@mail.uni-kiel.de
ak@informatik.uni-kiel.de
[2] Institute of Agricultural Engineering
Kiel University, Germany
amelfsen@ilv.uni-kiel.de

Abstract. Disciplines like life and natural sciences could gain high benefits from process mining in terms of identifying anomalies in the process or supporting predictive analytics in what is being measured. These disciplines, however, mostly work with data at a much lower level of abstraction and the data does not directly relate to high-level business process concepts as required for process mining. This paper discusses an approach for process mining on video data. As a use case, we applied our approach on video surveillance data of pigpens. Although, our process analytics pipeline from raw video data to a discovered process model has not yet been fully implemented, we are convinced that our approach is an essential contribution towards a (semi)automatic technique aiming to replace manual work.

Keywords: process mining · activity recognition · video labeling.

1 Introduction

Process mining is an established technique to give insights into data in terms of a structured order of activities (i.e., a process model). In this way, process mining allows identifying bottlenecks or compliance issues in business events. Mainly, process mining relies on business event data that is used as input to process mining algorithms and thus the data is expected to be on a high (business) abstraction level. Despite the success of process mining in the business context, process mining can provide an additional benefit to disciplines dealing with high volume and veracity of data. These disciplines like life or natural science have a high demand for a structured approach to answer process related questions like (1) what unknown processes are acting (i.e., did we find all processes that exist) and (2) whether the found processes actually work as thought.

Previously, we suggested approaches to discover process models from sensor event data [3] and "raw" time series data [12] with the purpose to give new insights

* Joined project with the Institute of Agricultural Engineering at Kiel University. Thank you for providing the data set and use case.

J. Manner, D. Lübke, S. Haarmann, S. Kolb, N. Herzberg, O. Kopp (Eds.): 14th ZEUS Workshop, ZEUS 2022, Bamberg, held virtually due to Covid-19 pandemic, Germany, 24–25 February 2022, published at http://ceur-ws.org

into the data in terms of the identification of anomalies in the process flow aiming to prevent unintended consequences. This paper presents our approach to discover process models from video data. As a use case, we applied our approach on video surveillance data of pigpens. So far, the behavior of pigs has been studied manually. Therefore, our approach aims to provide a (semi)automatic approach for pig behavior analysis in terms of health monitoring and understanding animal welfare. In this way, our approach makes a contribution to both questions (1) and (2) mentioned above.

The next section motivates why the use case is an appropriate starting point to develop techniques for process mining on video data.

2 Challenges

Compared to low-level raw data like sensor event data and time series used as input for activity recognition, video surveillance data of pigpens on the one hand eases the extraction of process activities, but on the other hand several challenges as mentioned below have to be overcome. Reasons facilitating the analysis are: (1) the behavior of pigs is limited to a few activities, which significantly simplifies activity detection compared to recognition of human activities in smart homes or smart factories. (2) A distinction between individual pigs is not necessary. This significantly simplifies the entity-centricity, which is challenging in smart homes where usually multiple objects are moving that need to be distinguished from each other.

To apply process mining on video data, however, requires bridging the following challenges: (1) no appropriate reference data set and labeled data exist. The freely accessible video-based data sets are mainly for object detection of other use cases like autonomous driving. Large computer vision libraries like Facebook AI Research's Detectron2 [10] grant access to trained neural networks, however, the detection of pigs is not covered by the commonly used COCO (Common Objects in Context) [5] and ImageNet [2] datasets. We found two pig-specific data sets for detecting positions and orientation [7] and tracking [8], but these data sets do not suit process discovery purposes. Almost no process-specific data exist in the data set. Therefore a high manual effort is required since neither labeled data nor an appropriate data analysis pipeline exist for our use case. (2) Image quality significantly correlates with the analysis results. Image quality is affected by image resolution, camera angle and camera quality. We initially received a data set of very low quality. In addition, the data set was not representative (i.e., too short image sequences). Therefore, recording of a new data set was necessary with a camera installation from a different angle. (3) Image noise (e.g., due to randomly switching from day to night mode, camera pollution and distortion due to neighboring pigpens). Finally, we recorded a new representative data set of higher image quality and less image noise.

The next section presents our approach aiming to address the challenges mentioned above.

3 Approach

Figure 1 shows our approach to discover a process model from video data with the following sequential steps:

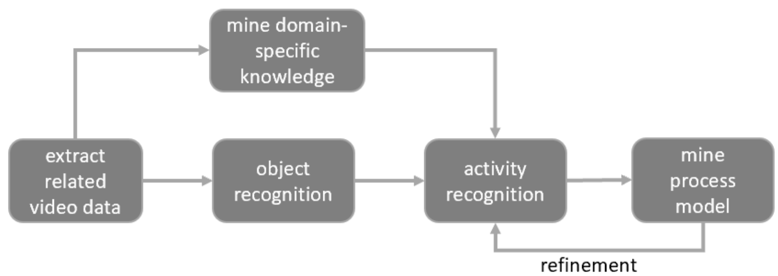


Fig. 1: The steps from video data to process discovery.

- *extract related video data from original data set*: we observed four pigpens, each with ten to twelve pigs, over a period of a few weeks. We recorded video material with a resolution of 1920x1080 pixels and 12.5 fps every day from 6:00 a.m. to 6:00 p.m. Mostly, the pig behavior does not change. Instead the pigs are in a kind of dormant phase. Many interesting actions only take place over a very short period of time, sometimes lasting just a few seconds. To detect related actions in our data set, we developed an algorithm measuring the movement intensity of a video sequence, which makes it easy to recognize the active phases of the pigs (see Figure 2). A spike in the chart indicates a new action.
- *mine domain-specific knowledge*: in this step we aim to identify context-related information that enhances action and object recognition. For instance, the location of the movement areas varies from groups of pigs. A group of young pigs would divide the pigpen differently than a group of older pigs. Thus, context information in terms of pig specificity is necessary in order to not distort the analysis results. Although, multiple data mining techniques have been used to mine domain-specific knowledge, again no specific technique exists for our use case. Therefore, the techniques have been tailored to our use case. First, we aimed to identify areas of high (visual) actions. The algorithm presented before has been enhanced to identify active movement areas. In general, a pigpen is divided into these three areas: sleeping/resting area, defecation area and feeding area. To automatically detect these areas, we used a slightly modified version of our motion intensity detection algorithm. We divided the images of a video into an area of 20x20 tiles and calculated the intensity of each tile over the entire video. Next, we converted the results into a 20x20 heatmap and easily identified the active areas. Figure 3 shows

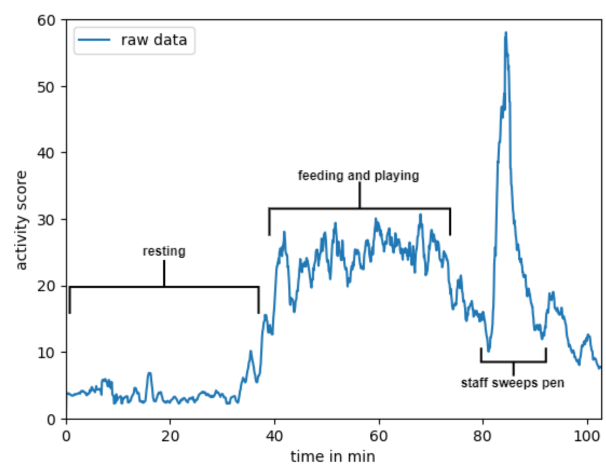
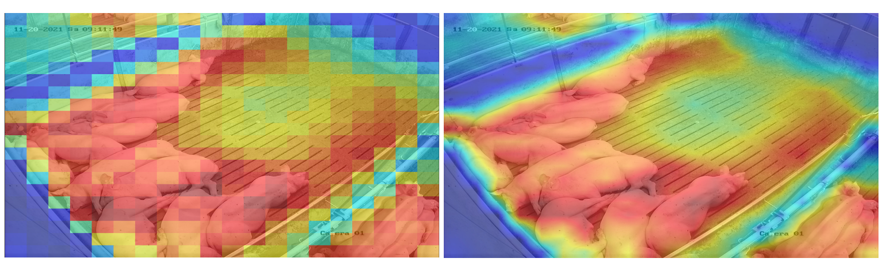


Fig. 2: Example output of our algorithm to determine movement intensity in a video sequence.

an example. Then, knowledge of the positions of all pigs over time is used to create a heatmap of common pig positions. To do this, we calculate the midpoint of each bounding box detected on the video. The position heatmap is then constructed from the relative frequency of midpoints per heatmap bin (see Figure 4 for a log-normalized example output of this analysis).
Tracking traces have been clustered to find common movement patterns (and paths between common areas). Figure 5 shows an example of 150 movement trajectories extracted from one video. Different movement patterns can be observed, e.g. the pigs are mostly stationary in their resting area.



(a) Original movement intensity heatmap of a video divided into 20x20 rectangles.

(b) Smoothed movement intensity heatmap of the original version.

Fig. 3: Example of our algorithm to explore the movement intensity of areas in the video.

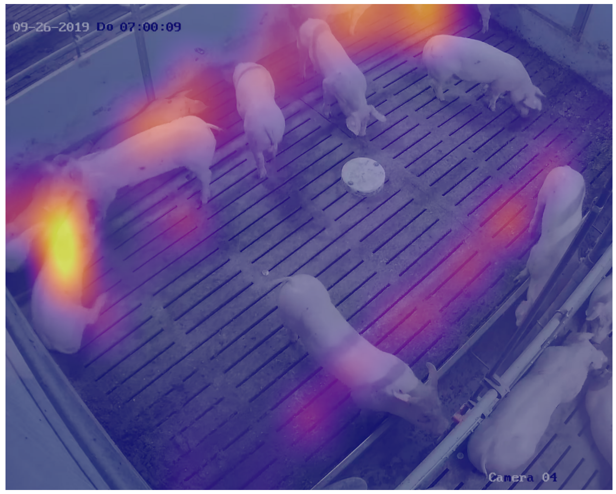


Fig. 4: Example output of our algorithm to determine common positions in a video sequence.

- *object recognition*: in this step, the video data is prepared for further analysis. First, an object detection is applied on the video. We chose YOLOv5 [4] due to its ease of use and ability to produce appropriate results with a relatively small amount of hand-labeled training data. Based on the object detection, (multi) object tracking has been used. This allows to analyze the same pig over multiple frames. We chose the DeepSORT (Deep Simple and Online Realtime Tracking) [9] algorithm to implement the tracking. DeepSORT is a well-established algorithm. The algorithm has been shown to work in a similar context to ours [1] and performs reasonably well on our data set without additional training. In the future, improved solutions for object detection and tracking could be applied to improve the quality of tracking results. However, many other solutions for the multiple object tracking problem require labeled tracking data for training. Since we aim to reduce manual labeling effort, the implementation of other tracking algorithms should be in proportion to manual effort. While a tracking dataset is available for pigs [8], it does not match our camera setup exactly. Also, there is no any labeled tracking data available when applying the analysis process in a different domain. If it was on purpose, the tracking results could be even used to localize individual pigs [1].
- *recognize activities in video*: The prepared video sequences and the associated position data from the tracking can be used as input for activity detection. In this step, also a model to learn visual features could be used. The learning process would have to be designed in a way where the features correspond to low-level events of the underlying process of the video. These low-level events can then be used to create event logs. While several techniques for

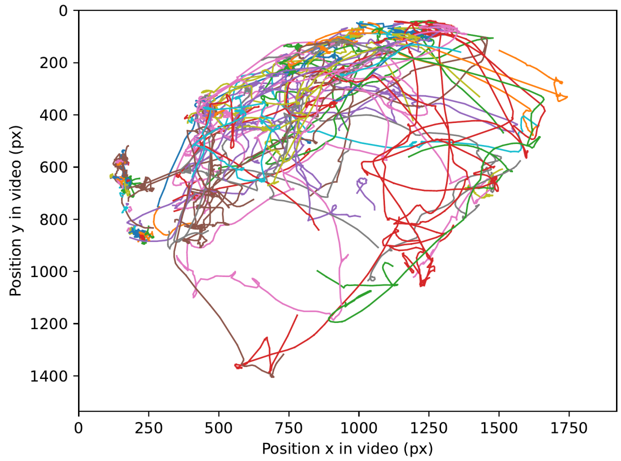


Fig. 5: Example output of our algorithm to track pig movements in a video sequence. In this example, 150 movement trajectories are shown.

pig activity recognition exist [11], they are either very specific to the unique properties of pigs or very specific to one type of activity (i.e. lying, standing, aggression). We choose not to use pig-specific techniques in this step to keep the approach generic.

- *discover process model*: the activities from the last step need to be aggregated/abstracted and enhanced with domain-specific knowledge (see step 2). Then, a case ID has to be created, e.g., according to the movement areas. A process model can then be mined from the event log.
- *refinement*: use the quality of the resulting process model to optimize the activity recognition and process model discovery.

4 Summary and Outlook

In studies of agricultural science alterations in behavior processes of pigs can be a helpful tool for analyzing and evaluating animal behavior, animal health and environmental impact. However, most approaches on identifying pig behavior based on video data only focus on single activities like e.g., lying, eating without analyzing the process. This paper suggested a process mining-pipeline to extract a process model from video data. As a use case, we applied our approach on video surveillance data of pigpens. Beside animal health, welfare and thermal comfort state, our approach can be used as a helpful indicator to evaluate and adjust climate conditions in mechanically ventilated barns. Likewise observations of activity and feed intake, which will vary depending on different climate conditions, supports the control of the above [6]. We see further use cases for our approach in medicine and material science that also handle large volume and veracity of data. Our approach of process mining on video data might be in medicine and material science for predictive analytics and outlier detection, which we believe to be more challenging than the current use case. Both assumptions that facilitate the analysis (i.e., low number of activities and entity-centricity) need to be bridged for an efficient solution.

References

1. Cowton, J., Kyriazakis, I., Bacardit, J.: Automated individual pig localisation, tracking and behaviour metric extraction using deep learning. IEEE Access **7**, 108049–108060 (2019)
2. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
3. Janssen, D., Mannhardt, F., Koschmider, A., van Zelst, S.J.: Process model discovery from sensor event data. In: Process Mining Workshops - ICPM 2020 International Workshops, Padua, Italy, October 5-8, 2020. LNBIP, vol. 406, pp. 69–81. Springer (2020)
4. Jocher, G., Stoken, A., Chaurasia, A., Borovec, J., NanoCode012, TaoXie, et al.: ultralytics/yolov5: v6.0 - YOLOv5n 'Nano' models, Roboflow integration, TensorFlow export, OpenCV DNN support (Oct 2021). https://doi.org/10.5281/zenodo.5563715, `https://doi.org/10.5281/zenodo.5563715`
5. Lin, T., Maire, M., Belongie, S.J., Bourdev, L.D., Girshick, R.B., Hays, J., et al.: Microsoft COCO: common objects in context. CoRR **abs/1405.0312** (2014), `http://arxiv.org/abs/1405.0312`
6. Nasirahmadi, A., Hensel, O., Edwards, S., Sturm, B.: A new approach for categorizing pig lying behaviour based on a delaunay triangulation method. Animal **11**(1), 131–139 (2017). https://doi.org/https://doi.org/10.1017/S1751731116001208
7. Psota, E.T., Mittek, M., Pérez, L.C., Schmidt, T., Mote, B.: Multi-pig part detection and association with a fully-convolutional network. Sensors **19**(4), 852 (2019)
8. Psota, E.T., Schmidt, T., Mote, B., Pérez, L.C.: Long-term tracking of group-housed livestock using keypoint detection and map estimation for individual animal identification. Sensors **20**(13), 3670 (2020)
9. Wojke, N., Bewley, A., Paulus, D.: Simple online and realtime tracking with a deep association metric. In: 2017 IEEE international conference on image processing (ICIP). pp. 3645–3649. IEEE (2017)
10. Wu, Y., Kirillov, A., Massa, F., Lo, W.Y., Girshick, R.: Detectron2. `https://github.com/facebookresearch/detectron2` (2019)
11. Yang, Q., Xiao, D.: A review of video-based pig behavior recognition. Applied Animal Behaviour Science
12. Ziolkowski, T., Koschmider, A., Schubert, R., Renz, M.: Process mining for time series data. Technical report (2022)