TURKEY BEHAVIOR IDENTIFICATION USING VIDEO ANALYTICS AND OBJECT TRACKING

Shengtai Ju[1], Marisa A. Erasmus[2], Fengqing Zhu[1], and Amy R. Reibman[1]

[1]School of Electrical and Computer Engineering, [2]Department of Animal Sciences,
Purdue University, West Lafayette, Indiana, USA

ABSTRACT

In this paper, we propose a method to identify behavior of experimental turkeys by automatically analyzing video recordings. Monitoring turkey health during production is crucial for improved turkey production. Turkey health can be reflected through their common behavior, and changes in the frequency and duration of their behavior can be used to detect sick turkeys early. Video recordings can be manually annotated to assist identifying turkey behaviors, but this is both time consuming and labor intensive. In this paper, we monitor and detect changes in turkey behavior using video analytics. Behaviors of interest include eating, drinking, preening, and pecking. Identifying these behaviors requires accurate estimates of turkeys' and turkey heads' locations. Reidentification of each turkey is crucial after significant shape deformation such as wing flapping and fast walking. Therefore, our system integrates a state-of-the-art turkey tracker and a head tracker with a behavior identification module to identify turkey behavior. Results demonstrate that our system is effective and accurate at estimating the spatial location of turkeys and their heads, and identifying all behaviors of interest with high recall.

***Index Terms*—** Video Analytics, Multi-Object Tracking, Animal Welfare

1. INTRODUCTION

Turkey production is important to the poultry meat industry, both in the United States and worldwide. During the production stage, turkeys are threatened by many factors including disease and poor environmental conditions. Turkey health can be reflected through their common behavior, and is directly related to turkey production since unhealthy turkeys are not optimal for production [1]. Sick and healthy turkeys exhibit distinct behavior, and early detection of sick turkeys through behavior analysis is critical.

One common approach used by animal scientists is to use an accelerometer mounted on each animal [2] to detect changes in walking behavior. An RFID (radio-frequency identification) system has been applied to detect eating and drinking behavior in pigs [3], but this only considers location-based behaviors. Video monitoring is a third option for analyzing animal behavior. Video has the advantage of being non-intrusive and only requiring one sensor for a room full of birds. However, a major limitation is that animal researchers must spend many hours annotating videos, which is fatiguing and error-prone.

Therefore, in this paper, we introduce a system that applies video analytics to automatically track turkeys and identify their behavior. Relevant behaviors associated with turkey health include: eating, drinking, walking, preening, and aggressive interactions such as beak pecking [4]. Preening is the action of a turkey using its own beak to manipulate its own feathers on the wings, back or breast. Beak pecking is the action of one turkey forcefully pecking at another turkey's head or face. Behaviors like eating and drinking can be detected by estimating the turkeys' or turkey heads' proximity to the feeder and drinker. For behaviors like preening and pecking, movement in the head regions are critical to finding these behaviors. Without accurate estimates of the location of the turkey heads, it is difficult to identify the turkey behaviors of interest. Therefore, we need both a reliable turkey tracker and a reliable turkey head tracker.

Many recently proposed object trackers focus on tracking pedestrians or focus on designing a single tracker effective for a wide range of applications [5]. Here, we focus on designing and implementing a tracker to solve a specific practical agricultural problem, where we incorporate specific domain knowledge about body shape and movement into our tracker. In addition, our work considers the implications that the performance of one object tracker may have a significant effect on the performance of tracking a related object. Specifically, we also consider important interactions between the accuracy of the turkey tracker and that of the head tracker.

Fig. 1. Experimental room with seven turkeys.

Tracking turkeys in a confined environment can be challenging in several ways. Figure 1 shows a frame from our turkey videos. As can be seen, the white turkeys look highly similar. Even with the markings that have been added to assist human viewers, the turkeys are difficult to distinguish, especially when occluded. Turkeys can also be highly-deformable during actions such as wing flapping and high-speed walking. They can rotate in different angles and expand their wings to many times their normal body sizes quickly. After these deformations, it is important for the tracker to re-identify each turkey so that we can continuously monitor each turkey's behavior. Furthermore, the turkey heads are relatively small compared to the size of the bodies, which makes tracking the heads challenging. Another challenging factor is the background color, which is quite similar to the turkey heads. These factors make our task of accurately estimating turkey head locations more challenging.

In [6], we introduced a comprehensive turkey tracking system, which includes a turkey tracker, a head tracker, a behavior identification module, and a graphical user interface that provides interactivity for researchers. Here, we consider the individual components. We detail the turkey and head trackers and their performance, and add the challenging behaviors of preening and pecking to the eating, drinking, and walking detection that we considered in [6]. Our turkey tracker relies on DeepSort [7], which performs significantly better for our application than the CSRDCF tracker [8] we explored in [9]; however, both trackers require modifications to become sufficiently accurate to facilitate acceptable behavior identification in our application.

In Section 2 we describe our overall system and provide details on the individual components. Due to the challenges presented by tracking a turkey head, the head tracker relies on the deep-learning turkey tracker to create an accurate localization. A specific behavior identification module is needed because no other system detects the types of behavior we are interested in. Tracking and behavior identification results are demonstrated in Section 3, while Section 4 presents concluding thoughts.

2. TURKEY BEHAVIOR IDENTIFICATION SYSTEM

2.1. System Overview

Figure 2 shows the overall block diagram of our turkey model. The turkey tracker is used to estimate each turkey's location in each frame. The head tracker further uses the turkeys' location information and estimate each turkey head's location. The behavior identification modules uses the estimated spatial location of both turkeys and their heads to predict turkey behavior. Optical flow is also used in the behavior identification module to detect movement of turkeys. Details of each component follow.

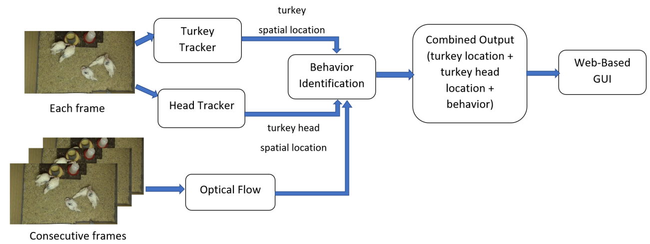

Fig. 2. Overview of proposed system.

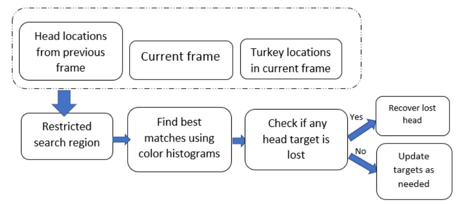

Fig. 3. Head tracker details.

2.2. Turkey Tracker

Our turkey tracker must simultaneously track multiple nearly identical objects while maintaining their identities, a task known as multiple object tracking (MOT) [5]. Correlation filter trackers such as the Kernelized Correlation Filter Tracker (KCF) [10], and Discriminative Correlation Filter Tracker with Channel and Spatial Reliability (CSRDCF) [8], have demonstrated good tracking performance [11]. Recent approaches take advantage of the recent advancement in object detection and computer vision to create detection-based deep-learning trackers. DeepSort [7], for example, combines Kalman filtering [12] and deep neural networks to perform multi-object tracking. It uses an object detector to generate bounding boxes and updates the Kalman filters based on the detection results.

In this paper, our turkey tracker is largely based on the DeepSort tracker [7] using YOLOv3 [13] as the object detector. Detection results are generated on each frame of the video and used as measurements to update the Kalman filter. The Kalman filter estimates the spatial location of each turkey using the available measurements and associates each newly detected bounding box to an existing track using the Mahalanobis distance. For long-term association, feature representations are used to re-identify each turkey and assign a unique ID to each. Re-identification is also done after significant turkey deformation such as wing flapping and high-speed walking. For feature extraction, a convolutional neural network (CNN), trained on the MARS dataset [14], is used. Cosine distance is calculated in the feature space to match each new detection to an existing tracked object. Compared to [9], we added Kalman filtering to predict the locations of turkeys based on past measurements. We also added a feature descriptor to extract turkey features to maintain their identities.

Simply implementing [7] did not work well on our data; therefore, we made a few changes to the original implementation. Compared to the original implementation in [7] and for the Kalman filters, we only track the position of the objects and ignore the aspect ratio since turkeys rotate their bodies. We have also increased the number of successful and consecutive detection necessary to initialize a new track. Our enhanced turkey tracker allows us to build a reliable head tracker and perform subsequent behavior analysis.

2.3. Head Tracker

Accurate estimates of head locations are critical so that we can identify turkey behaviors that indicate changes in turkey welfare. As mentioned in the introduction, turkey heads are difficult to track because of their sizes, the lack of distinctive RGB information, and the background color. Therefore, we implement a pre-processing step to make the turkey heads appear more distinguishable from the background during head-tracking, by multiplying the S channel from the *HSV* color space by a constant. We use color histograms to detect possible head patches in every frame and associate each detection temporally by computing the nearest neighbors.

The head tracker is illustrated in Figure 3. A restricted search region is formed by considering the head locations in the previous frame and the turkey locations in the current frame. After defining our search region for heads, we find the best-matched patches using color histograms and overlapping windows. In addition, color histograms of the turkey heads are updated to account for any appearance changes. We describe the details of the head tracker in the following paragraphs.

Because the appearance of turkey heads vary in different videos due to lighting variations and camera angles, we manually initialize the turkey head locations at the beginning of the video with a 25×25 bounding box. The head tacker uses color features extracted from the *HSV* and *CIELAB* color space to search for best match in subsequent frames. The best match is defined as the image patch having the smallest cosine distance in the feature space with the target patch. However, whether the identified best match is actually a good estimate of the real head location remains unknown. To solve this issue, we check if any of the turkey heads are lost during tracking. A head tracker is considered to be lost if either of two conditions occur: 1) the predicted turkey head is more than 150 pixels away from its body, 2) the cosine distance between the color histograms of the target template and the identified best match is larger than 1.3 in the feature space. These thresholds are determined based on empirical observation. To reinitialize a lost head tracker, we search for the best-matching square patch within the corresponding turkey bounding box.

Another issue when tracking a small template over time is the need to update the target information. As the video progresses, our targets of interest, the turkey heads, could have moved to locations where the lighting condition has changed. Then, our manually initiated turkey heads would no longer be accurate representations of the turkey heads in the current frame. To overcome this issue, we need to periodically update the head targets throughout the video. Every ninety frames, we assess whether the head targets are still good representations of the turkey heads in the current frame. The reason for this is because if we update the targets too often, error can accumulate after each update. If we update the targets with error each time, eventually we will have inaccurate representations of our targets. Here, error comes from the best match drifting away from the actual head center and capturing more and more background. Therefore, we check whether the current targets allow us to find heads with relatively small cosine distance error in the current frame. If not, we update the head targets using the best matches in the current frame. The frequency of this assessment is determined ad hoc based on the characteristics of the video.

2.4. Behavior Identification Module

This section presents the behavior identification module, which relies on the spatial location estimated by the turkey tracker and head tracker. The behaviors of interest can be classified into two groups: location-based behaviors, and motion-based behaviors. Walking, eating, and drinking can be categorized as location-based behaviors. Walking is identified by computing the distance traveled by each turkey within a fixed time period. Eating and drinking can be identified by computing the distance between each turkey's head and the feeder and the drinker. In [6], we considered only location-based behaviors, whereas both location-based and motion-based behaviors are considered in this work.

For the motion-based behaviors, spatial locations are used along with optical flow. For preening and beak pecking, accurate estimates of the head locations and detection of movement in the head regions are necessary. Preening is detected when a turkey's head is close to its body and movement is detected in a sequence of frames. For beak pecking, the first step is to identify two close turkey heads. If their heads are close enough, motion detected from optical flow is used to determine whether there is pecking or not, based on an empirically set threshold. The magnitude of optical flow vectors is summed in the head regions and compared to the threshold.

3. EXPERIMENTS AND RESULTS

3.1. Dataset

Our methods are evaluated on a set of videos that consist of white commercial turkeys kept in experimental pens at the Purdue College of Veterinary Medicine. Videos are captured using consumer grade cameras at 1280×720 (HD) resolution with 30 frames per second (FPS). As shown in Figure 1, food and water are available to the turkeys and turkeys have similar sizes. The YOLOv3 detector is trained using 490 turkey images with seven turkeys in the experimental room. We evaluate our methods on two 3-minute video clips and six 1-minute video clips, with seven turkeys in the experimental room.

3.2. Evaluation of Trackers

We use multi-object tracking accuracy (MOTA) and multi-object tracking precision (MOTP) [15] as the evaluation metrics to assess the turkey and the head trackers as shown in Equations 1 and 2, respectively. m_t, fp_t, and mme_t are the number of misses, the number of false positives, and the number of mismatches at time t, respectively. g_t is the number of objects present at time t. d_t^i is the distance between a matched object in the ground truth with its corresponding hypothesis, at time t. c_t is the number of matches, at time t. MOTA increases with improved performance while MOTP decreases. MOTA is reported in Table 1 as percentage and MOTP is reported as pixels per match. Since our goal is to detect changes in turkey behavior, percentage and numbers on commonly-used MOT metrics are not the only criteria for evaluation. Trackers can also be evaluated by the effectiveness of the behavior identification results and by how much valuable information researchers can extract from the tracking results.

$$MOTA = 1 - \frac{\sum_t (m_t + fp_t + mme_t)}{\sum_t g_t} \tag{1}$$

$$MOTP = \frac{\sum_{i,t} d_t^i}{\sum_t c_t} \tag{2}$$

Clip Length	Method	Turkey Tracker		Head Tracker	
		MOTA%	MOTP	MOTA%	MOTP
3-min	[9]	−59.6	32.0	−69.4	26.1
	Ours	**70.1**	**20.1**	**26.0**	**25.8**
3-min	[9]	10.7	21.6	−50.4	29.5
	Ours	**91.6**	**13.9**	**63.2**	**17.8**
1-min	[9]	−25.4	33.5	−20.1	**24.0**
	Ours	**76.1**	**19.9**	**41.0**	25.2
1-min	[9]	−14.9	25.8	−41.0	28.7
	Ours	**79.0**	**18.8**	**37.0**	**25.3**
1-min	[9]	−37.2	27.8	−16.7	26.4
	Ours	**55.6**	**21.7**	**26.2**	**24.8**
1-min	[9]	12.3	20.5	−15.3	27.3
	Ours	**91.3**	**12.5**	**70.7**	**19.4**
1-min	[9]	40.3	21.4	26.6	23.4
	Ours	**93.9**	**16.0**	**50.2**	**16.1**
1-min	[9]	5.6	28.5	16.7	21.2
	Ours	**89.6**	**13.9**	**66.6**	**17.6**

Table 1. Turkey tracking and turkey heads tracking results. (Bold numbers are best results for each clip and each task.)

Table 1 shows the results of tracker evaluation. As can be seen from the results, our proposed method in this paper outperforms our previous tracker [9] by a significant amount, in both short-term and long-term tracking. Due to its high turkey tracking accuracy, the new tracker also shows promising results in tracking turkey heads. Regarding turkey re-identification after highly-deformable actions, there are a total of six instances of wing flapping and fast walking in our two 3-min evaluation clips. Our proposed tracker successfully re-identified the turkeys in five of six instances, whereas [9] and [7] only succeeded in two of six instances.

3.3. Analysis of Behavior Identification

To evaluate our behavior identification module, we use the activity recognition metrics including: precision, recall, number of insertion, and number of deletion [16]. An insertion (**I**) is defined as an event from the system's output with no corresponding ground truth event. A deletion (**D**) is defined as an event that the system completely fails to detect.

Behavior	Precision %	Recall %	# of I	# of D
Walking	62.5	90.0	15	2
Eating	73.0	85.0	3	1
Drinking	33.3	100.0	3	1
Preening	45.0	100.0	11	0
Beak Pecking	61.5	89.0	5	1

Table 2. Behavior identification results.

Table 2 shows the results of the turkey behavior identification module evaluation. The results show that our system can detect each behavior well, as shown by the high recall. The number of deletion (**D**) is small which means we are not missing many events. For our application, recall is more important than precision because we want to detect as many events as possible and lower the number of missing detection. The high recall and low number of deletion across all behaviors demonstrate that our turkey tracker and head tracker can successfully detect turkey behavior and provide useful information to researchers.

4. CONCLUSIONS AND FUTURE WORK

In this paper, we propose a system to detect turkey behavior using object tracking and video analytics. Moreover, a DeepSort-based turkey tracker is implemented. A novel head tracker using color histograms is used to estimate the spatial locations of turkey heads. A behavior identification module is built to identify different behaviors of turkeys including: eating, drinking, walking, preening, and beak pecking. Our model achieves high tracking accuracy in both tasks of turkey tracking and turkey head tracking. It also shows promising results in detecting turkey behaviors. Future research will consider adding more types of behavior into the behavior identification module and applying our method to a commercial farm to test its robustness under more challenging conditions.

5. REFERENCES

[1] N. van Staaveren, E. M. Leishman, B. J. Wood, A. Harlander-Matauschek, and C. F. Baes, “Farmers’ perceptions about health and welfare issues in turkey production,” *Frontiers in veterinary science*, vol. 7, 2020.

[2] R. Stevenson, H. A. Dalton, and M. Erasmus, “Validity of micro-data loggers to determine walking activity of turkeys and effects on turkey gait,” *Frontiers in veterinary science*, vol. 5, p. 319, 2019.

[3] S. G. Matthews, A. L. Miller, J. Clapp, T. Plötz, and I. Kyriazakis, “Early detection of health and welfare compromises through automated detection of behavioural changes in pigs,” *The Veterinary Journal*, vol. 217, pp. 43–51, 2016.

[4] M. A. Erasmus, “Welfare issues in turkey production,” *Advances in Poultry Welfare*, pp. 263–291, 2018.

[5] W. Luo, J. Xing, A. Milan, X. Zhang, W. Liu, and T.-K. Kim, “Multiple object tracking: A literature review,” *Artificial Intelligence*, p. 103448, 2020.

[6] S. Ju, S. Mahapatra, M. A. Erasmus, A. R. Reibman, and F. Zhu, “Turkey behavior identification system with a GUI using deep learning and video analytics,” *IS&T Electronic Imaging*, 2021.

[7] N. Wojke, A. Bewley, and D. Paulus, “Simple online and realtime tracking with a deep association metric,” *IEEE International Conference on Image Processing*, pp. 3645–3649, 2017.

[8] A. Lukežič, T. Vojir, L. Čehovin Zajc, J. Matas, and M. Kristan, “Discriminative correlation filter with channel and spatial reliability,” *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[9] S. Ju, M. A. Erasmus, A. R. Reibman, and F. Zhu, “Video tracking to monitor turkey welfare,” *IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI)*, 2020.

[10] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, “High-speed tracking with kernelized correlation filters,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 583–596, 2014.

[11] M. Fiaz, A. Mahmood, and S. K. Jung, “Tracking noisy targets: A review of recent object tracking approaches,” *arXiv preprint arXiv:1802.03098*, 2018.

[12] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *Transactions of the ASME–Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.

[13] J. Redmon and A. Farhadi, “Yolov3: An incremental improvement,” *arXiv preprint arXiv:1804.02767*, 2018.

[14] L. Zheng, Z. Bie, Y. Sun, J. Wang, C. Su, S. Wang, and Q. Tian, “Mars: A video benchmark for large-scale person re-identification,” *European Conference on Computer Vision (ECCV)*, 2016.

[15] K. Bernardin and R. Stiefelhagen, “Evaluating multiple object tracking performance: the clear mot metrics,” *EURASIP Journal on Image and Video Processing*, vol. 2008, pp. 1–10, 2008.

[16] J. A. Ward, P. Lukowicz, and H. W. Gellersen, “Performance metrics for activity recognition,” *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 1, pp. 1–23, 2011.