A New Ectotherm 3D Tracking and Behavior Analytics System Using a Depth-based Approach with Color Validation, with Preliminary Data on Kihansi Spray Toad (Nectophrynoides asperginis) Activity

Philip Bal
Damian Lyons
Avishai Shuter

Follow this and additional works at: https://fordham.bepress.com/frcv_facultypubs

Part of the Artificial Intelligence and Robotics Commons, Behavior and Ethology Commons, and the Robotics Commons
A New Ectotherm 3D Tracking and Behavior Analytics System Using a Depth-based Approach with Color Validation, with Preliminary Data on Kihansi Spray Toad (*Nectophrynoides asperginis*) Activity

**PHILIP BAL**
Fordham University, Computer & Information Science, JMH 340, 441 E. Fordham Rd, Bronx, New York 10458, USA

**DAMIAN LYONS**
Fordham University, Computer & Information Science, JMH 340, 441 E. Fordham Rd, Bronx, New York 10458, USA

**AVISHAI D. SHUTER**
Department of Herpetology, Wildlife Conservation Society’s Bronx Zoo, 2300 Southern Boulevard, Bronx, New York, USA

Automated target tracking has been an active research area for many years, and varied approaches have been developed in tracking aerial, land-based, and underwater targets (Bar-Shalom and Fortmann 1988). Multitarget point tracking, multisensor image fusion, and using visual data to track multi-part targets were all developed to solve unique problems related to tracking objects in 3D-space (Reid 1978; Rasmussen and Hager 1998; Sharma 1999). While these technologies have been applied to limited use in animals, tracking human targets from visual or RGB-D (combined depth and color image, e.g., Microsoft Kinect cameras) has received the majority of research attention (Teixeira et al. 2010).


The core components of modern animal tracking systems depend primarily on the specific use for which they are designed. To track wild animals in their natural habitats, implanted or attached GPS loggers are often used due to their record of effectiveness (Schofield et al. 2007). Camera traps have proven effective for surveying the presence of smaller varieties of both mammals and ectotherms. Using a HALT (optical light) trigger, Hobbs and Brehme (2017) surpassed the effectiveness of traditional passive infrared (PIR) camera traps (often used for medium or large mammals). Passive infrared cameras typically need around 2.7°C temperature difference between an animal and its surroundings for activation. This presents a difficulty for use with ectotherms, where body temperatures may not differ significantly from their environment.

HALT systems depend on a pre-aligned near infrared beam (NIR) that is elevated above ground level detritus. A major drawback of this approach, however, is the direct manipulation of the animal’s environment, as well as the need for the elevated platform to be free of debris to maximize effectiveness. Furthermore, camera traps are limited in scope to viewing areas, rather than tracking specific individuals or movements. Video cameras have also been used to track and analyze the behavior of wildlife; image analysis has been used to classify avian observations according to species. For example, pigeon behaviors, such as the ‘head-bobbing’ and the ‘foot-plants’ components of courtship, have been monitored from motion capture data and automatic image recognition criteria (Song et al. 2008; Zeiler et al. 2009).

Imaging devices that yield depth, as well as color information, at each pixel in an image (RGB-D cameras) have also been used in identifying and tracking animals. Mittal et al. (2017) focus on areas in the wilderness where human-elephant interaction is likely to occur and result in injury, and in this case, the presence of target animals is confirmed with a trained neural network using depth information to identify regions of movement in video frames. Similar depth-oriented approaches work effectively in enclosed spaces as well. Barnard et al. (2016) focus on structural classification of dog postures in shelters derived from depth information acquired from Microsoft’s Kinect cameras. Lyons et al. (2015) developed an algorithm to identify and count pigeon behaviors such as pecking and feeding. While there are similar tracking approaches and applications that exist for both humans and animals, there are relatively few implementations for smaller animals.

Gomez-Martin et al. (2012) described a comprehensive computer vision package, Sensory Orientation Software (SOS), for automated measurements of animal posture and movement. Recordings were made using a Stingray Camera (Allied Vision Technologies) equipped with a 12–36 mm, 1:2.8, 2/3 C Computar lens (CBC Group). However, the authors commented on the sensitivity to disturbances of camera pose during measurements. EthoVision (Noldus Information Technology) is a program commercially developed initially for top-down 2D analog video tracking of mice in laboratory settings, with additional functionality available, such as a 3D tracking option (Noldus et al. 2001). However, the software’s primary laboratory use case is inherently a small, clearly defined region of interest with no visual obstructions and would thus not work as well in environments containing foliage, less than well-defined regions of interest, arbitrary camera positions and angles, or with target subjects similar in color to their surroundings. Another commercial software, AnTracks (AnTracks Computer Vision Systems 2013) tracks individual ants using color information but lacks any 3D tracking options and has limited behavioral analytic outputs. Open source options, such as BioSense, are available as well, though they primarily focus on 2D tracking, based solely on color information, and lack significant behavior analytics tools (Patman et al. 2018).

The tracking system presented herein, the Ectotherm Tracking Algorithm (ETA), allows an experimenter to set up the RGB-D camera at a convenient viewing location and generates multi-target 3D tracking information from depth-change and color correlation measurements for even small, fast moving targets. Additional programs analyze the 3D target tracks to present them in a manner useful for animal behavior researchers (Shneiderman and Plaisant 2019). This system was developed through trials using Kihansi Spray Toads (Nectophrynoides asperginis) at the Wildlife Conservation Society’s Bronx Zoo, New York, USA as tracking targets.

The Kihansi Spray Toad, declared extinct in the wild by the International Union for Conservation of Nature (IUCN) in 2005 (IUCN SSC Amphibian Specialist Group 2015), is a toad native to the mist zone of the Kihansi River waterfall in the Eastern Arc Mountains of Tanzania, Africa (Lee et al. 2006; Fig. 1). The Bronx Zoo propagates these toads as part of an ongoing project with the aim of reestablishing a self-sustaining population of the species in its native habitat (Lee et al. 2006). Initial efforts at breeding the toads were successful, and thousands have been reintroduced into the wild, with thousands more still being bred at the Bronx Zoo (Lee et al. 2006; Shuter, pers. obs.). Given the size of the Kihansi Spray Toad population currently housed at the Bronx Zoo, the opportunity presents itself to study the complex behaviors of these animals at scale for the first time.

However, due to the impracticality of manually observing thousands of toads for extended periods of time, an automated solution was required. Traditional automated tracking systems use radio-based telemetry, which depend on manual insertion of a sensor in or on a target animal, or camera traps that use infrared technology in order to identify potential targets from body heat. Because these approaches only work for animals that are endothermic or large enough to have a sensor attached, automated tracking systems for diminutive, ectothermic animals are more difficult to design. To our knowledge, there does not exist a cost-effective automated tracking system for multiple small, ectothermic animals that is both able to track targets in 3D space and which does not involve direct interaction with the specimen(s) in question. However, it is important to note that prototypes of camera trap systems for small ectotherms are being experimented with, though they cannot track the movements of specimens (Hobbs and Brehme 2017). Herein, we propose a new tracking system in order to automate the tracking of small ectothermic animals which will be of general interest to behaviorists, behavioral ecologists, behavioral neuroscientists, ethologists, evolutionary ecologists, and others interested in automated behavior logging for small ectotherms.

**Materials and Methods**

The Kihansi Spray Toad captive breeding program consists of a collection of over 60 terrariums housing more than 2000 toads in environments that closely mimic their natural habitat (Channing et al. 2006; Shuter et al. 2016). Tracking toads that average less than 3 cm in size using a video camera is difficult due to the small size of the animals and their tendency to blend...
into the background. We employed an Intel RealSense SR300 series camera (Intel Corporation) to collect both depth and color information. Leveraging the fine depth detail that the SR300 camera provides, image regions in which depth changes were recorded over time were tagged as candidate toad targets. When this initial target selection is further filtered with a toad target color correlation operation, it becomes possible to identify a toad’s discrete movement with high confidence. The tracking program operates at frame rates using a standard, retail computer. This tracking system records multiple target movement tracks to log files in 15-minute intervals. The log files are then processed and displayed in a variety of formats to support the identification of gross activity patterns, high and low-density movement areas in an enclosure, and more. As very little information exists concerning the activity levels and social behaviors of Kihansi Spray Toads, and the resources to collect data are limited, the proposed tracking system provides a passive and inexpensive way to collect large amounts of novel data. Moreover, the proposed tracking system has potential to be extended, with relative ease, to any species of small ectothermic animal, thus providing a completely unique and novel tracking solution.

Ectotherm Tracking Algorithm.—To begin logging tracks, the Intel RealSense SR300 camera was positioned approximately 2.5 cm away from the glass-walled terrarium (Fig. 2a). The video stream, consisting of depth and color streams, was captured at 30 FPS at an image resolution of 640 × 480. The depth stream was set for close range using Intel’s preset. All tracking and data analysis programs were run on an Intel NUC (Next Unit of Computing) with a 7th generation 1.9 GHz Intel i5 processor, 8GB of DDR4 2400 MT/s RAM, and a 5TB external hard drive for storage of the recorded streams (Fig. 2b). Video stream recording and access used Intel’s RealSense 2.0 SDK version 2.15.0. Further processing of video frames was done using operators from OpenCV version 3.3.0.

Contour Identification.—The first step in target tracking is the identification of image regions corresponding to potential targets in the current depth and color image. In our approach, targets were identified from depth-based movement information and then validated using color correlation (Fig. 5).

To obtain movement information in the form of contours from the depth stream, a foreground depth mask was generated using a Gaussian Mixture based background subtraction model (KaewTraKulPong and Bowden 2002). Let $I_f$ represent an image from the RGB-D camera stream at frame $t$. Furthermore, let $I_f$ be separable into $I_f^b$ and $I_f^d$ containing the RGB image, and the depth image, respectively. Background subtraction ($BS$) was applied to $I_f^d$ to extract a foreground representing depth changes:

$$BS(I_f^d) = (B_f, F_f)$$

where $B_f$ is the background and $F_f$ is the foreground image. $F_f$ was then converted to binary and processed using morphological operations to improve the target contour for tracking as shown in Equation 2, where $Erode^2 = Erode(Erode())$. An example of such a binary, processed foreground is in Fig. 3a.

$$F_f^b = Blur^3(Dilate^6(Erode^2(Binary(F_f))))$$

The contour location function is represented as function $FC$ with $\delta_i$ representing the $i$th contour,

$$FC(F_f) = \{\delta_i | i \in 0...max_{contours} and A_{min} < \delta_i area < A_{max}\}$$

$FC$’s output does not include any contour that that is greater in area than constant $A_{max}$ or less in area than constant $A_{min}$ so that movements that are either too small or too large to be a toad were not tracked. This means all $\delta_i$ satisfy $A_{min} < \delta_i area < A_{max}$ s.t. $\delta_i area$ is a contour’s area.

The visual and depth components of the visual image were linearly aligned so that color and depth values at a point in the scene can be related. Constants $X_{offset}, Y_{offset}$, and $Y_{offset}$ transformed any pixel coordinates $(u,v)$ as shown in Equation 4.
The common region of interest (ROI) shared by both cameras is shown in Fig. 3b, the depth image, and Fig. 3c, the corresponding RGB image. As the depth stream from the camera has a larger field of view than the color stream, the ROI is represented by a bounding rectangle in Fig. 3b.

While it was possible to start tracking with only the contours derived from depth information, it was important to consider how to minimize false toad track identifications. The terrariums that the toads live in were exposed to dynamic lighting conditions over the course of the day, as well as active misting systems (to both maintain adequate moisture/humidity levels and replicate their native environment) that formed water droplets on leaves that caused them to move. This movement ultimately resulted in erroneous depth contours. To minimize erroneous tracking of these none-toad contours, color correlation was used to achieve a higher confidence that a toad’s movement created the contour identified in the depth image.

The dynamic lighting conditions that the toads were exposed to were accounted for by converting the RGB images to the HSV (hue, saturation, value) color model. However, $I_f^c$ was converted to HSV format with the value component set to a constant, $v_v$, so that light levels did not influence color correlation, thus $HSV(I_f^c) = (h, s, v_v)$, as illustrated in Fig. 4a.

Because the ROI did not encompass all contours derived from the depth frame, frame bounds needed to be established. As color correlation was performed, contours were tested to fall in bounds of the color frame. Bounds on the depth frame were defined by $x_{min}, x_{max}, y_{min},$ and $y_{max}$. Contours were mapped from the depth image to RGB image using $I_f^d(u, v) = (d)$. Any contour that did not satisfy the following conditions was eliminated: $x_{min} \leq x \leq x_{max}$ and $y_{min} \leq y \leq y_{max}$.

Color correlation (CC) was performed using a pre-selected HSV color template of a toad, $I_{toad}$: Equation 5. The output of the color correlation was a matrix of mean squared difference of pixel values, $I_{cc}$, e.g., Fig. 4b.

$$CC(I_{toad}) = I_{cc}$$

To determine whether the correlation match was significant, we confirmed that the minimum value of the color correlation was below threshold constant $d_{cc}$. The location of the correlation minimum was $P_{cc}$.

$$P_{cc} = \arg\min_{u, v} I_{cc}(u, v)$$

If $I_{cc}(u_{cc}, v_{cc}) \geq d_{cc}$, $\delta_t$ was rejected for further processing, as the color was not close enough to that of the toad template. The depth contour was rejected as a false target measurement.

Color correlation could not be performed at night, when color values cannot be determined. Furthermore, a light source could not be added as this may have both disturbed the toads and/or altered their normal behaviors. However, the depth information was still present. To sense if light levels fall below an acceptable level for color correlation, a constant $v_{light}$ was defined. The pixel values of the value channel, $v$, of HSV image
before they are set to 1, were added cumulatively to yield a value $V_{\text{total}}$ and checked if they were less than $V_{\text{light}}$. In any frame where $V_{\text{total}} < V_{\text{light}}$, the color correlation step was skipped. This allowed tracking to continue in complete darkness. However, without color correlation, tracking at night could be more error prone.

**Tracking.**—In order to track individual targets, the contour information from each successive image $I_j$ needed to be related to contours in prior images. The motion of toads as recorded in the camera stream was converted for each target observed to a sequence of points in space, $P_l = (x, y, z), l \in \mathbb{N}$, that correspond to the location of the toad in successive images in the camera stream. Determining which contour in an image is the best match for each track constructed from prior images is the Data Association problem (Tian et al. 2018). The tracking literature has developed many approaches to Data Association to handle, e.g., associating the correct contours after tracks appear to occlude and cross one another from the camera viewpoint. Because we had access to the contour locations in depth, the target crossing problem was greatly simplified. For that reason, we adopted a fast and straightforward approach to data association: nearest neighbor matching. This association procedure is described in Fig. 6.

Let $m_k \in M_j \ s.t. M_j$ contain all of the color-validated contour measurements in the common ROI of frame $j$ with $m_i$ as the $ith$ valid contour’s measurements.

\[ l(t_k) = P_k, \hspace{0.5cm} i = 1, 2, \ldots, N \]

To perform nearest neighbor tracking the following attributes were defined. Let $t_k \in T_j$ s.t. $T_j$ contains all of the tracks up to, but not including, frame $j$, and $t_k$ is the $ith$ track. $h(t_k) = (P_0, P_1, P_2, \ldots, P_l)$ are the points (attributes) associated with track $t_k$. $P_0$ to $P_l$ represent the attributes of a track in $l$ consecutive frames. An individual track attribute, $P_i$ is described as $P_i = (x, y, z)$ the centroid of the validated contour matched to the target in that frame. The function $c(m_j) = (x, y, z)$ was applied to any contour $m_j$ to yield its centroid. Let $l(t_k) = P_k$, the last point in the track so far, and $f(t_k) = l$, the number of points in the track so far.

The distance from the last known location of a track, $l(t_k) = P_k$, to every valid contour $m$ centroid was calculated as a $|T_j| \times |M_j|$ matrix, where element $a_{k,j}$ was defined as:

\[ a_{k,j} = ||l(t_k) - c(m_j)|| \]
A minimum distance assignment of measurements to tracks was calculated using the Kuhn-Munkres (Hungarian) algorithm, where each track is extended by the region that yields the best overall solution. However, the distance \( d_k \) was then checked for each track that \( d_k < d_{\text{min}} \) s.t. \( d_{\text{min}} \) is a constant threshold distance. If the condition did not pass, \( t_k \) was noted as not having appeared in frame \( j \) and its current position, \( P_{ts+1} \), was repeated as \( P_t \). If the condition passes, then the track, \( t_k \) was associated with its nearest neighbor contour \( m_j \).

Position filtering, to account for noise in the region measurements, was performed for track \( t_k \) as follows: 
\[
P_{(t_k)_{ts+1}} = \text{filtered}(m_j, t_k)
\]
provided that the track length \( f(t_k) > r_j \) s.t. \( r_j \) is a constant that describes the minimum threshold number of frames that a track has existed for. If the comparison was not satisfied, filtering does not occur. If the comparison was satisfied, filtering was performed over the time window \( P_{(t_k)_{ts+1}} \)...

**Behavioral Data Analytics.**—All tracks were collected over 15 minute intervals and logged to uniquely named files that were then processed to generate behavioral analytics for researchers studying animal behavior. In this study, the analytics described include 3D and 2D track graphs, 2D heatmaps, gross track meeting graphs (potential fighting and potential mating), gross activity graphs, and gross activity comparison graphs.

**3D Track Graphs.**—The lowest level track information was presented by a 3D plot used to visualize the tracks identified by the ETA. Each track, \( t_k \), was plotted in a 3D coordinate plane as a colored line with different tracks in different colors, viewable from any perspective (Fig. 7a). However, it may be difficult to piece together behaviors from the raw track information: there may be a lot of tracks and the relationship and synchrony between them may not be immediately obvious. For this reason, additional track and activity visualizations and analytics were developed.

**2D Track Graphs.**—The 3D track information was easily transformed so as to generate a synthetic view as if the viewer were positioned in front of the terrarium, or if the viewer were looking from above the terrarium. For the front view graph, only the \((x, y)\) coordinates of \( t_k \) are plotted as described in Equation 8.

\[
h(t_k^x) = (P_0^{xy}, \ldots, P_l^{xy}) \text{ s.t. } P_i^{xy} = (x, y)
\]

where \( (P_i = (x, y, z)) \in h(t_k) \) for the top view graph only the \((x, z)\) coordinates were plotted, in a similar manner to above.

As with the 3D graph, each track \( t_k \) in the 2D graphs was represented by a separate line. The colors of each track will be consistent across the 3D and 2D graphs. As seen in Figs. 7b (the front view graph) and 7c (the top view graph), there is a black bounding box containing the tracks. In the front view (Fig. 7b) this bounding box corresponds to the ROI seen in the depth image of Fig. 3b. For the top view (Fig. 7c), only the x boundaries correspond to the ROI seen in Fig. 3b, as the depth values had no constraints on which objects are tracked.

**2D Heatmaps.**—2D heatmaps were generated in order to more easily identify spatial areas of higher and lower activity. The front view heat map divides the horizontal x coordinate into \( H_x \) bins (where \( H_x = 5 \) in our examples), and the vertical y coordinate into \( H_y \) bins (where \( H_y = 5 \) also in our examples). Each coarse cell in the heatmap, \( H_q \) where \( q \in \{1 \ldots H_x\} \times \{1 \ldots H_y\} \) shows the spatial activity in the region of space viewed by that portion of the image as a pseudocolor. For each track \( t_k \), the relevant coordinates (either \((x, y)\) for front view, or \((x, z)\) for top view) were checked to see which cell of the heatmap they fall in. Let \( R(H_q) \) be the region of the plane covered by cell \( H_q \). The heatmap cell, \( H_q \), was incremented by 1 for every unique track that falls in it. If a track had multiple points in a cell, or enters and exits a cell any number of times, \( H_q \) was still only incremented by 1, thus giving each \( H_q \) the exact value of unique tracks that pass through it, either in part or their entirety. Of course, if a track passes through multiple cells, each individual \( H_q \) that the track passes through was incremented. Once all \( T \) tracks are processed, intervals of numbers of unique tracks were established and depicted via pseudocolor in the heatmaps. Equation 9 shows the condensed logic for each cell \( H_q \) for the front and top view heatmaps.

\[
H_q = \sum_{i=1}^{T} g h(i, q), gh(i, q) = \begin{cases} 1, & \text{at most once if } t_i \cap R(H_q) \neq \emptyset \\ 0, & \text{otherwise} \end{cases}
\]

Similarly to the 2D graphs, the front view heatmap(s) (as seen in Fig. 8a) and top view heatmap(s) (as seen in Fig. 8b) have bounding rectangles that correspond to the depth ROI as seen in Fig. 3b.

**Gross Activity.**—The gross activity graphs show general activity trends over the course of multiple hours or days in 15-minute intervals. The number of unique tracks present at each interval \( T_m \) was calculated from the timestamp associated with each recorded track coordinate. In all of the following

---

**Fig. 7.** Tracks generated over the course of the same 1-h interval, with each track represented by a separate line in space. A) 3D \((x, y, z)\) rendering of tracks; B) 2D rendering of front view \((x, y)\) of tracks; C) 2D rendering of top ("bird’s-eye") view \((x, z)\) of tracks.
graphs, all $T_m$ are 15-minute intervals, with points plotted at the lower bound of the interval to represent the gross number of target behaviors found in the entirety of the interval. To better compare gross activity trends over the course of a day, graphs of separate days were overlaid with each line segment representing a different day, the entire work flow of which is described by Equation 10 where $T$ is the total number of tracks.

\[
G(T_m) = \sum_{i=1}^{T} \sum_{j=1}^{m} g_i(t, T_m), \quad g_i(t, T_m) = \begin{cases} 
1, & \text{if } t_j \text{ in interval } T_m \\
0, & \text{otherwise}
\end{cases}
\] (10)

**Gross Track Meetings.**—Activity trends of toads moving towards or away from one another were examined by calculating the gross number of unique track meetings, where a meeting is defined as 2 toads within a constant distance $d_{cr}$ of one another at the same time. The gross number of unique meetings that fall in a time interval was then calculated as is summarized in Equation 11, where $r$ represents a timestamp, where $T$ is the total number of tracks, and where we define for convenience $t_x(r) = P_x = (x,y,z)$.

**Gross Track Mating Meetings.**—The toad meeting measurements of the previous section reflect measurements of any time that toads approached one another, whether the toads had any behavioral goal in the proximity or not. As an example of how processing of track information can be customized by some a priori knowledge of animal behavior, we further refined general meetings into potential meetings with the behavioral goal of mating and potential meetings with the behavioral goal of engaging in combat (e.g., defense of territory). We did this based on looking at the track information just before and just after the meeting event. The objective with this measurement was to give a behavioral researcher a shortcut to determining when mating behaviors are potentially happening, which can then be backed up and validated by visual inspection of the video data.

In order to determine if tracks represented a mating pattern by toads being tracked, the software tracked instances where two tracks approach each other and then merged into one, and continued moving. We established a mating threshold distance, $d_m = 0.25 * d_{cr}$, and we searched for any 2 tracks that got within that distance, where one track stops, and the other continues. These steps are summarized in Equation 12 with function "Matings" calculating all potential mating meetings for the interval $T_m$, where $\Delta$ is a small fixed time interval. As there may have been multiple toads attempting to mate at the same time, and the same toad may have attempted to mate with multiple toads during a single track’s duration, no further constraints were placed on what constitutes a mating meeting.

\[
Matings(T_m) = \sum_{i=1}^{T} \sum_{j=1}^{m} S(r, T_m)W(r, i, j, d_m)A(r + \Delta, i, j)
\] (12)

\[
A(r, i, j) = \begin{cases} 
1, & \text{if } t_i(r) = \emptyset \land t_j(r) \neq \emptyset \lor (t_i(r) \neq \emptyset \land t_j(r) = \emptyset) \\
0, & \text{otherwise}
\end{cases}
\]

**Gross Track Fight Meetings.**—We also classified track meetings into a “potential fight” category. When engaged in territorial disputes, male toads will typically move within a certain distance of one another, $d_f = 0.5 * d_{cr}$, and then proceed to move away from each other. Furthermore, only two toads typically fight at a given time (Shuter, pers. obs.). For simplicity, we condensed this notion to a 1-track following rule such that each track may have had only one “fight” with one other track for its entire duration. Equation 13 summarizes these steps with function “Fights” calculating all potential fights meetings for time intervals, $T_m$.

**Gross Track Scaling.**—Since lighting conditions at night were too dim for color correlation, the overall number of tracks measured at night were larger than that during the day, due to
noise in the raw depth data. We made the assumption that this noise is approximately similar at all times and we estimated its magnitude during the day, when we could use color correlation to validate targets from the raw depth data. To account for this daytime/nighttime variation in tracking in the gross activity graphs, we used the daytime data (between 0500 h and 1900 h) to calculate an estimate of the noise, yielding a conversation ratio. All the tracks that were confirmed with color correlation were divided by the number of all the potential tracks that are observed (regardless if they had been confirmed by color correlation or not), then the gross number of tracks at each interval during the night (where no color correlation is performed) was scaled by the conversion ratio to account for the effect of increased noise at nighttime.

The enclosures were lightly misted with filtered water at regular intervals via a custom, automated system in order to simulate the conditions of the toads’ natural habitat, however this misting also caused an increase in erroneous target identifications due to the increase in contours generated from moving leaves and water droplets (Shuter et al. 2016). The misting apparatus was on a timer, and hence the times during which misting will affect tracking could be predicted. To account for this, manual random samples of misting and non-misting video segments were observed to calculate how often tracks were correctly identified as toads during respective misting and non-misting periods. Each of the data points in gross activity graphs were then multiplied by either the misting or non-misting ratio depending on which time interval they fell in. The resulting graphs are therefore normalized to account for any uncontrollable external factors.

\[
Fights(T_m) = \sum_{i=1}^{T-1} \sum_{j=i+1}^{T} S(r, T_m) W(r, i, j, d) L(r + \Delta, i, j, d)
\]

\[
L(r, i, j, d) = \begin{cases} 
1, & ||t_i(r), t_j(r)|| > d \\
0, & \text{otherwise}
\end{cases}
\] (13)

RESULTS

Final output graphs of the ETA system were collected over approximately 2 days and consist of 3D and 2D track graphs, 2D heatmaps, gross track meeting graphs (potential fighting and potential mating), gross activity graphs, and gross activity comparison graphs. 3D and 2D track graphs, as well as 2D heatmaps, were discretely generated for each 15-minute interval of recorded footage for ease of interpretation, while cumulative gross graphs were generated over the full length of time of recorded footage. It is noteworthy that the tank filmed was being prepared for shipment and thus had more toads than a typical tank, which might have had an impact on the animals’ behavior.

DISCUSSION

In this study we have presented a tracking and behavioral analytics system for automated observation of the Kihansi Spray Toad. Our work was motivated by the opportunity to study the toad, declared extinct in the wild, at the Bronx Zoo where thousands of the animals are being bred for reintroduction to the wild. Because continual direct human observation of toad behavior is challenging, an automated tracking and behavioral analytics system is described and developed in this paper. Challenges to automated tracking include the quantity and size of toads as well as the fact that they are ectothermic. Historically, captive animal ethogram studies have employed a number of data collection strategies including direct observation at random or predetermined intervals (Lee et al. 2006; Rija et al. 2014), post hoc review of video recordings (Shuter and Sardelis 2018), and strategically placed passive transponder-reading units (Bauert et al. 2007). Current understanding of Kihansi Spray Toad behavior is based on direct observations of both wild toads and those under human care. Initial observations about the natural history of this species were based on direct observation of animals while they could still be found in the wild, and Channing et al. (2006) reported that the toads are diurnal, that males are territorial towards other males, and suggested that different parts of the habitat were utilized by toads of different age groups. The same study also postulated that habitat use was determined by the time of day and that individual toads maintained “home ranges,” both of which could potentially be confirmed by the ETA.

Following the extinction of the species in the wild, subsequent studies made observations based on animals kept under human care. Multiple researchers have independently observed the territorial behavior of male toads towards their male rivals (Lee et al. 2006; Arch et al. 2011; Rija et al. 2014; Shuter and Sardelis 2018). Despite these reports, questions remain about the limits of male Kihansi Spray Toad territories—both spatially and temporally—which could ultimately be answered by utilization of the ETA system. While previous studies illuminated important aspects of the Kihansi Spray Toad behavioral repertoire, and even aided in the formation of successful ex situ assurance colonies that ultimately served as the source of individuals that
were reintroduced into the wild, many questions about the finer details of the species’ behavior remain.

As of now, the most comprehensive detailing of Kihansi Spray Toad daily activity patterns and behavior was presented by Rija et al. (2014). This study, performed at two Tanzanian facilities that house the toads, recorded direct observations made by researchers for 20 minutes at a time per study enclosure between 0700–0930 h and 1600–1830 h over the course of several weeks (Rija et al. 2014). Some key findings of this study included that the toads observed spent significantly more time “resting” than any of the other behaviors being recorded, that toads seemed to be more active in the evening, and that toads spent more time on logs and plants than on the ground (Rija et al. 2014).

Our preliminary findings can be added to the pool of data collected by these previous studies. Gross activity levels in the study enclosure over the study period show a clear pattern that is congruent with previous observations (Fig. 9). The ETA reveals that toads are most active while the automated mist cycle is activated, and that while the animals are indeed diurnal, they are also active in the evenings. The objective of the behavior analytics output is to provide behavior researchers with a shortcut to identifying portions of stored video that yield fruitful behavioral insights. Two example analytics were those for measuring gross activity in potential toad matings and in potential toad fights. Moreover, analytics generated for a two-day period suggest that useful spatial and temporal findings may both be confirmed and expanded on pending more rigorous investigation.

There are certain shortcomings in the current hardware and software. When a subject is particularly close to the Intel RealSense SR300 camera, there is a non-negligible degree of visual disparity that causes the depth-derived contours and RGB images to align incorrectly, thus producing erroneous identifications, as may be seen in Fig. 13. However, the small size of the toads required the camera to be relatively close to the terrarium. A more sophisticated alignment procedure, or improved camera hardware could address this issue.

As the ETA system was designed with specifically small ectotherms in mind the effective tracking area is relatively small, thus requiring either a controlled environment or small home range of the animal being tracked. Implementation of this technique in the field is possible provided moderate foresight is given to weatherproofing the required hardware and providing an

![Fig. 11. The total gross potential mating meetings graph generated over a course of 3 days for a total of approximately 48 h.](image1)

![Fig. 12. The total gross potential fighting meetings graph generated over a course of 3 days for a total of approximately 48 h.](image2)

![Fig. 13. An example of the close range visual disparity issue where it can be observed, by the manually drawn red boxes, that the RGB location of toad (to the left of the large plant) does not match up with the sensed depth location of the toad (in front of the large plant).](image3)
ample power supply. Moreover, tracking in the wild necessitates animals be directed into the line of sight of the camera, by means such as a drift fence, or by placing the camera directly adjacent to a small, static, target species’ habitat.

Our primary objective has been to develop working and useful behavior analytics, so as to demonstrate the potential for animal behavior specialists. However, certain aspects of the measurement, tracking, and analytics can be improved. Color correlation, while fast, is a relatively simple approach to validating targets. Given the progress in neural network-based image recognition, we believe that would yield superior recognition results without sacrificing computational performance. Moreover, machine learning algorithms in general have been shown to produce insights into animal behavior that may be useful in producing novel hypothesis. Finally, the behavior analytics implemented here should be seen as a first iteration of what is necessary, and we expect to modify and improve these as the system is used by animal behavior experts.

Acknowledgments.—We are grateful to the government of the United Republic of Tanzania for their ongoing efforts to protect the Kihansi Spray Toad, and we offer thanks to our Tanzanian colleagues, whose research endeavors are invaluable to our own. We gratefully acknowledge the support of the WCS IACUC Board, as well as the support given to this project by WCS Bronx Zoo leadership and staff, whose research endeavors are invaluable to our own. We gratefully acknowledge the support of the WCS IACUC Board, as well as the support given to this project by WCS Bronx Zoo leadership and staff, including Don Boyer, Kevin Torregrosa, Micah Siegel, and Elana Frank. We also thank Armando Califano for developing the first prototype of the toad tracking system, as well as Fordham University for providing funding for supplies.

Literature Cited