State Estimation of a Wild Robot Toward Validation of Rigid Body Simulation

James R. Taylor[1] and Evan Drumwright[2]

***Abstract*— There exist few objective measures to evaluate or compare multi-rigid body dynamics simulators. This absence creates uncertainty in simulation capabilities and accuracy. Simulation science has used theory and other simulations (*verification*) and real-world data (*validation*) to evaluate simulation correctness. With respect to rigid body dynamics, ballistic rigid body motion has been *validated*, but simulations involving contact and friction frequently seem to produce results that appear inconsistent with real-world observations; accurate validation has been seldom performed for contacting "rigid" bodies, likely because the observation problem is so challenging (compared to, e.g., fluid dynamics). This paper concentrates on a simple validation scenario for multi-rigid body dynamics with contact and friction, which are essential for simulating robotic locomotion and manipulation. We study collection and estimation of motion data from a mechanically simple but highly dynamic, real-world robot whose motion is primarily driven by contact and friction.**

I. INTRODUCTION

Simulation supports design, prototyping, testing, and evolutionary stages of development long before production of any robot; however, confidence among the robotics community in multi-body dynamics simulation remains low. While dynamics simulation of ballistic rigid body motion is well known to be generally consistent with natural behavior, simulations involving contact and friction are prone to producing physically implausible results. Given that the contact and friction models have been validated by mathematicians, physicists, and engineers on simple systems over years of study, one must wonder where the software that implements these models fails. Indeed, the past decade has seen the continual release of new simulation software libraries (including one by the second author), each of which promises to eliminate the artifacts but in hindsight has improved the status quo little, if any.

In the past, multi-rigid body simulators have been judged based on performance against a prescribed set of scenarios (see, e.g., [3]) with each scenario designed to evaluate potential modeling errors. These scenarios are a good start, but likely not sufficient; for example, knowledge of multi-body dynamics simulation internals tells us that determining the degree of slope that a box on a ramp begins to slide is unlikely to point to the cause of robotic grasping artifacts.

We believe that those efforts that test elementary aspects of the physical models do remain useful. Nevertheless, this paper focuses on using *validation* (comparison against real world data) rather than *verification* (checking that the software implementation matches theory). Toward that effort, this paper studies and captures data from a *wild*, real-world robot. We provide inertial and visual models of the robot, a guide to capturing motion data from the robot, an initial data set, and state estimates of the real robot. To accomplish this goal, we developed a motion capture process relying on a VICON motion capture system, an inertial model of the robot using SOLIDWORKS, and a visualization of the motion capture data, and we carried out and published data from a set of motion capture sessions.

> Videos rendered using state of the art simulation software (GAZEBO using ODE) will show that this simple system, which should be readily modelable with multi-rigid body dynamics, displays far different behavior from that observed in simulation.

(a) (b) (c)

Fig. 1: (a) An off-the-shelf WEAZELBALL with the two hemispheres, equatorial o-ring, and central axis switch visible. (b) Interior view of the "free" hemisphere (left-yellow) and the "fixed" hemisphere (right-red). (c) "Forward" velocity v develops by friction between shell and environment due to angular velocity ω_s, a response to the torque generating pendulum angular velocity ω_p.

The subject is a spherical robot called a WEAZELBALL (WB) — Figure 1 — which meets a variety of desirable criteria: wide availability, low cost, simple mechanism, complex (wild) behavior. It has been the subject in a number of research efforts focused on controlling aggregate behavior of "wild bodies" [1] [2] [4]. The WB basic principle and design was patented more than a century ago [5], operates by a combination counter-torque drive and offset center-of-mass, and is both underactuated and nonholonomic. A motor drives a pendulum around an axis fixed inside a spherical shell which produces torque in the pendulum and a response torque in the shell — Figure 1(c). In an unconstrained environment, friction between the shell and the environment causes the counter-torque to drive the robot "forward" with respect to the rotational motion of the pendulum and offset center-of-mass causes the robot to wobble from side-to-side. Upon collision with an obstacle, the rudimentary mechanism

[1] George Washington University, Washington, DC 20052, USA
[2] Toyota Research Institute, Palo Alto, CA 94306, USA

generates a complex range of motions reminiscent of piroetting, rocking, and tumbling.

We present a process for capturing WB behavior with the desire that the process and data archive might be used, studied and improved. We believe successful simulation of real-world WB behaviors would mark a step toward better robotic simulation, so we want to promote the WB as a simulator dynamics metric and to emphasize accurate study of the subject over the process of data collection. We search for ways to estimate and validate the data collected to build a solid foundation, but we encourage others to recreate and refine this work.

This article has the following structure: fundamental measurements are defined in Section II, a description of our process to capture WB state follows in Section III, an explanation of post-processing and analysis of the process follows in Section IV, and a conclusion in Section V.

II. MEASUREMENTS

Simulation validation relies on analyzing state data, so evolution of WB state over time must be measured as accurately as possible. An off-the-shelf WB is not designed for state measurement. The spherical shell has few features that can easily be tracked and shell opacity obscures the state of the pendulum. Any modification to the WB will affect dynamics. For example, adding instrumentation alters mass distribution which changes system inertia and adding features to the surface alters surface geometry and friction characteristics which modifies contact behavior. The WB design makes some modification necessary to measure shell and pendulum state, so modifications are selected to minimize changes to mass distribution and surface features and also minimized to maximize reproducability of the data gathering process.

A. *Modeling the* WB

The WB is composed of two principal components, a spherical, rigid external *shell* and an internal, weighted *motor assembly* that acts as the pendulum. The spherical shell has two molded plastic hemispheres which thread together and a groove at the equator to mask the threads and hold an o-ring. One hemisphere has a port through the pole and the other is solid with a steel pin cast into the shell at the pole. The weighted motor assembly is mounted to and rotates around the pin which generates both torque and novel inertial configurations. When the shell is closed, the rotational axis of the motor assembly aligns with port and pin. *Fixed hemisphere* refers to the hemisphere with the motor assembly mounted to it and *free hemisphere* refers to the detachable hemisphere containing the port.

An inertial and visualization model for the WB was developed in SOLIDWORKS by studying a disassembled, spare WB. The shell is modeled by a hollow sphere. The motor assembly is modeled by a set of unique shapes with different material densities. The motor assembly model is composed of three component models: housing, battery, and weight. The housing is a set of cast plastic hulls that encloses motor, switch, gearing, and battery. All housing parts other than the battery are modeled together with homogeneous density as a single housing part. The battery is standard AA and the weight is a cast-steel hollow wedge. Component models are combined into a single assembly and assigned a reference frame consistent with the motor assembly rotational axis. SOLIDWORKS was queried for assembly inertial tensor and center-of-mass and Collada files were generated for visualization and both are provided in the repository.

B. *Measuring shell state*

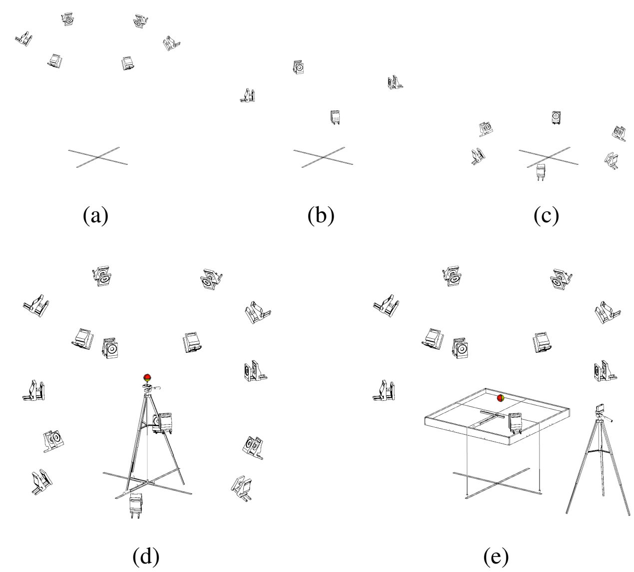


Fig. 2: VICON cameras are centered around an axes mark in three rings and focused on a point approximately 4ft from the floor. (a) 6 cameras form the high ring. (b) 4 cameras form the middle ring. (c) 6 cameras form the low ring. (d) The registration array includes the three camera rings and is assigned an origin aligned to the central axes. (e) The motion capture array includes the high and mid camera rings included in the registration array but excludes the low ring.

The VICON array used in this work consisted of sixteen motion capture cameras focused into a cube two feet above the ground and six feet on a side — Figure 2. In the center of this space, a large, orthogonal axis was marked on the floor. VICON cameras were set up in three rings with six mounted high-level, four mounted mid-level, and six mounted low-level. All three mounts were included in the registration array and high and mid-level mounts were included in the motion capture array. Cameras were posed to avoid dead spots in areas critical to registration and motion capture activities.

A WB left to freely roam will eventually explore an entire accessible planar environment; however, the need to focus the VICON system in a relatively small area constrains the collection environment and requires an enclosure. A square, inset table no larger than 4ft on a side and with rails no higher than 2in in depth was designed — Figure 3.

A standard, spherical marker is unsuitable for the WB as these markers alter contact behaviour by modifying friction, weight, and shell geometry. Flat retro-reflective tape which

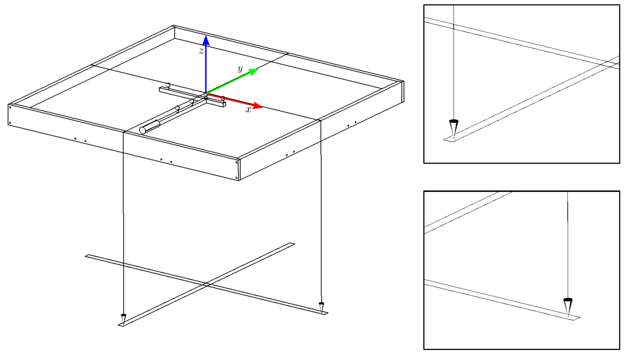
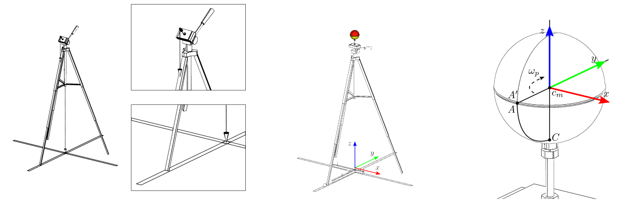


Fig. 3: The enclosure is centered to the capture array and allows the capture frame origin to be assigned to the enclosure center.



(a) (b) (c)

Fig. 4: The stand: (a) posing and aligning the stand to the registration array center, (b) assignment of the registration frame origin and a mounted WB, (c) the registration frame origin defines the orientation frame for the registered model, so the transform from the marked, mounted WB to a VICON pose is clearly defined.

slighly modifies contact parameters can be used without significantly affecting shell geometry.

Marker visibility is predominantly subject to occlusion events, *i.e.* where a marker is occluded from a camera by an opaque object. The more cameras a marker is visible to, the more accurate triangluation of that point. A VICON array works best with a marker visible to at least three cameras but will still triangulate a marker using two cameras with increased error. While the marker normal can be generalized as the surface normal, retro-reflective surfaces scatter light in many directions which allows light to return from retro-reflection back towards the emission point from many angles.

In lieu of spherical markers, two-dimensional circular markers are cut from retro-reflective tape and applied to the shell. Flat markers on the shell do modify friction, weight, and geometry, but by using a minimum number of markers and by keeping the marker size small effect on these parameters is minimized. Circular markers are applied to the surface of the shell in sufficient numbers such that the VICON array can track the shell using a large number of cameras. A marker applied to the surface of the shell forms a small dome which has less visibility than a three dimensional marker and a given dome is very often occluded to most cameras. Given a sufficient number and distribution of cameras and markers, the array can maintain continuous localization of the shell with acceptable error. When using dome markers, an array composed of a large number of cameras in a small area is necessary to ensure the subject remains visible to the array at all times.

Testing of our initial motion capture process revealed a significant amount of error from the motion tracking model which was controlled with a rigid stand to consistently pose the WB during registration — Figure 4. The stand enables VICON to focus on a fixed WB from many directions and marker application to a known WB orientation. The stand is a co-opted camera tripod with simple modifications to the mount plate to allow the head to be aligned with the registration frame and the WB to be aligned to the head.

C. Measuring motor assembly state

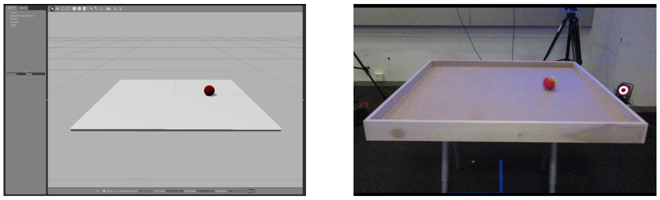
WB behavior is a product of counter-torque generated by the motor assembly, the rotating offset center-of-mass state w.r.t. shell orientation, contact and friction between shell and environment, and enviornmental collision and much of this behavior is due to the dynamic state of the hidden motor assembly. A number of ways to capture motor assembly state have been considered such as by adding an encoder or IMU, and casting a transparent shell; however, each of these alterations affect dynamic behavior by significantly changing inertia and friction.

A simple approach involves adding a basic, independent signal circuit that produces a visible signal each time the motor assembly completes a cycle. The circuit is powered by a 3V button-cell battery. The signaling circuit consists of the button-cell battery, a red LED, a resistor, and a small length of copper wire. The LED is mounted into the shell by boring a small hole through the shell at a dimple located near the coaxial push button. Light emitted by the LED is directly observable when the shell is viewed from almost any angle where the switch is visible. With this simple circuit, every cycle of the motor housing causes the battery leads to contact the LED leads which closes the circuit for a short but long enough period to clearly light the LED.

For each motion capture session, a HD CMOS camera records video of the WB at 30 fps to record signals from the motor assembly along with VICON motion captured shell state. During post-processing, each video is synchronized with VICON state data, each frame is scanned for LED signals, and if a signal is detected it is classified and recorded.

D. Visualization

GAZEBO helps visually review streamed VICON state using SOLIDWORKS Collada WB models. VICON state is read by a GAZEBO plugin to update a replay visualization of captured data. Visualization alone helps assess VICON array accuracy and allows capture instrument tuning, but visualization combined with frame-by-frame video analysis using **`melt`** allows for data synchronization, post-processing error correction, and movie generation — Figure 5.



(a) (b)

Fig. 5: Two frames from a synchronized visualization show WB state at a given time during motion capture: (a) The frame from GAZEBO visualization and (b) the frame from **`melt`**.

III. MOTION CAPTURE PROCESS

To plan WB motion capture is to plan both registration and collection activities. Camera placement must be planned, tested, and fine-tuned to optimally handle both activities, all instruments for all activities must be prepared, and transitions from one activity to another must be rehearsed. While visualization is a necessary component in post-processing, it also plays a critical role in setting up and fine-tuning WB marking and the motion capture array.

A. Model registration

Application of dome markers and WB registration must be systematic and controlled to maximize the number of markers visible to the array in all poses and to minimize error. To control marker application and validation and to control registration, the WB is fixed to the stand which is centered and aligned to the registration array — Figure 2(d). Once the array is ready, markers are added individually and systematically until no more markers can be applied.

The stand fixes a point on the WB equator to the registration frame and controls for shell orientation in the registration frame. The stand consists of a co-opted camera tripod and an additional plumb bob which is used to align both stand and head to the array — Figure 4(a). A mounting plate attaches WB to stand, aligns plate to the stand and registration frame, and provides alignment guides for the WB.

Once the shell is closed, a pen mark A is drawn across equator onto both hemispheres so they may be aligned along A in the future. Mark A is continued in an arc across the free hemisphere until it intersects the center of the port. Alignment mark A ensures the shell can be opened and closed for maintenance without need to reregister and aligns shell consistently to stand and registration frame. Once the WB is mounted to the stand — Figure 4(b) — mark A should be aligned to the mount — Figure 4(c). When marking A the position of the motor assembly should also be considered such that zero position will align the motor position with the zero position on the VICON model.

The low ring cameras were specifically added for registration of a fixed WB. Markers applied near the south pole are occluded to high cameras and may not be visible to enough mid cameras to be validated. In capture space, low cameras are occluded by the enclosure and low cameras create dead-spots for high and mid rings, so low cameras are excluded from the capture array; however, low cameras provides necessary coverage when registering the WB.

WB marking and registration requires adding a marker to the shell, registering the current marker configuration, then either correcting symmetry or adding another marker. Once a stable registration is found for the current configuration, another marker is added and the process is repeated. Using this process, eight markers were applied to the WB used in motion capture and attempts to add a ninth marker consistently resulted in symmetric errors in registration.

B. Motion capture

Motion capture involves setting the experimental table, switching from registration array to capture array, aligning the table, assigning the capture array origin, setting up the HD video camera, and streaming data from the capture array.

The experimental table consists of a small table as base, the enclosure top, and a number of shims to level the top. The table is centered and leveled with the array axes mark using centering notch plumb lines and shimming. The floor is checked for level at the enclosure center and in all four quadrants; the floor is leveled by inserting shims between the base and top until level across the enclosure floor. Taught line between the centering notches allows the wand to be aligned to one orientation in both registration and capture frames and to have an origin defined in the raised enclosure center. The low cameras are excluded and the remaining cameras are remasked. The VICON origin is set to the orientation defined by the wand in the enclosure — Figure 2(e).

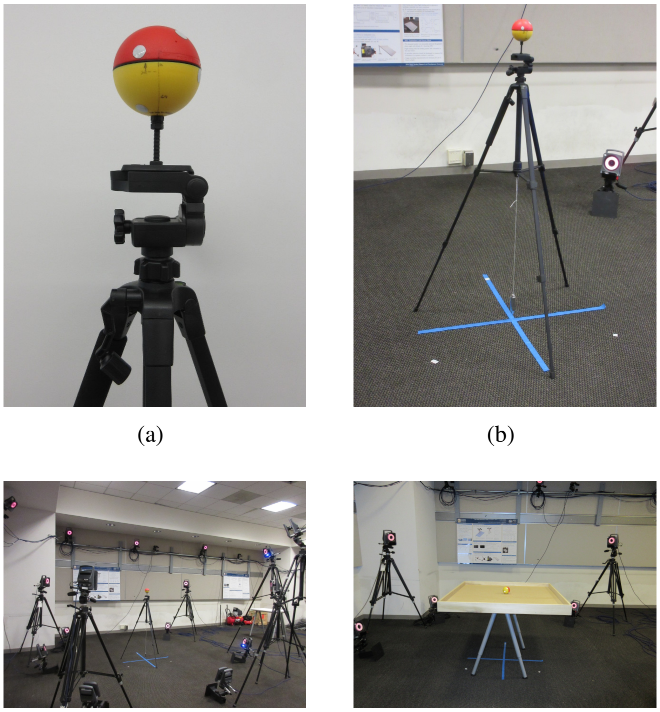
To capture the earliest possible signal on video, the WB is released into the enclosure with the LED visible to the video camera which is positioned to have a broad view of the enclosure. A general guideline would be to position the video camera perpendicularly to the typical WB release path with the free hemisphere facing the video camera.

C. Configuration validation

The array configuration is tested by rehearsing several collection test sessions with the WB on the table using all experimental apparati and activities, *i.e.* capturing the WB stream from VICON TRACKER, recording video of the experiment, and releasing the WB onto the table. Test session motion data is replayed using GAZEBO visualization — Section II-D — and marker placement quality and the incidence of particular array errors are evaluated. If there is significant incidence of error in WB orientation tracking, the table is removed from the array, the stand is replaced into the array, and marker registration is repeated.

D. Session collection

Each capture session is started by preparing motion capture software and video camera and by covering the WB by an opaque bag. TRACKER streams data whenever the model is in view, so the bag minimizes premature detection. Each session is started in the following order: WB is placed into bag, VRPN recording is started, video recording is started, WB is powered on, bag is placed into enclosure, WB is



(a) (b)

(c) (d)

Fig. 6: (a) Closeup of the marked WB mounted and aligned to the stand. (b) Stand aligned with the array origin. (c) View of the registration array. (d) View of the table in the capture array.

released from bag, bag is removed, and data is recorded for an alloted time. Once the session time has elapsed both VRPN and video recorders are stopped and all systems are reset for another session.

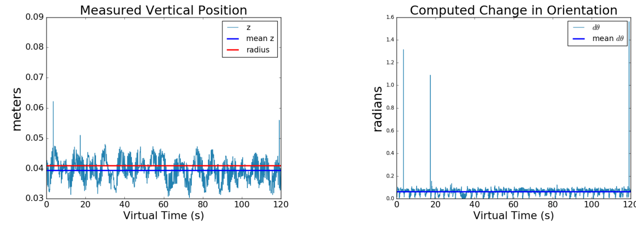
Ten total motion capture sessions were carried out to develop initial sample data. Each motion capture session lasted approximately two minutes and five seconds and was recorded using a 100 Hz VICON array and a 30Hz HD video camera. Sixteen cameras were used in the registration array and shared ten cameras with the capture array

IV. POST-PROCESSING

Once session data is collected, raw VICON and video data must be processed to identify error in the array, synchronize the data streams, and classify motor assembly signals.

A. VICON *analysis*

In vizualization, two anomalous behaviors were demonstrated in raw VICON data as sources of error, the visualized WB oscillates between interpenetrating and floating above the table and the visualized WB spontaneously jumps from the table with a simultaneous snap change in orientation. Analysis focused on vertical position z which should be equal to the WB radius — ideally $z = 41$mm — and on change in quaternion orientation $d\theta$ between the current frame and the next frame. Figure 7 shows that there is noise and position error in all measurements and that the spontaneous events are reflected in both z and $d\theta$ measurements.

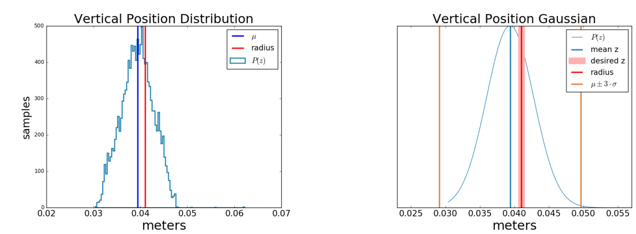


(a) (b)

Fig. 7: Plots of the first session are representative of all sessions. (a) Vertical position z should equal WB radius but raw data is noisy, mean should overlay radius but there is a clear difference, and spikes beyond noise suggest the WB jumps above the table at spontaneous samples. (b) Change in orientation between each sample exhibits similar error and spikes suggest the WB makes spontaneous snap orientation changes in the same samples as spontaneous jumps.

Spontaneous events are attributed to marker occlusion resulting in a loss of sufficient marker visibility to VICON. Event incidence is subject to camera and marker placement, so the number of spontaneous events in a session is a metric for the quality of the experimental setup. In a well-tuned array, spontaneous events are infrequent and occur primarily when the WB enters a corner where video confirms the WB stops, so highly infrequent spontaneous events can be interpolated.

The oscillation anomaly and the offset between mean and WB radius demonstrated in Figure 7(a) are partly attributed to the manual assignment of center in the tracking model. Manual assignment results in an unknown offset between the tracking model and the subject which requires estimation of center offset and compensation in VICON data.



(a) (b)

Fig. 8: (a) Vertical position histogram suggests a Gaussian with a mean left of radius while spikes in Figure 7 appear as right outliers. (b) Ideal vertical measurements would be tightly clustered around radius, but the shifted mean, wide bell, and existence of large outliers indicate various error in raw signal data.

Capture array accuracy is expressed in Figure 8 where properties of the gaussian reflect tracking model and array error which are reduced through fine tuning before session execution. Right outliers are spontaneous anomalies, difference between mean and radius tracks error in center assignment, and standard deviation reflects error in the detection system. Integrating generation of these plots into Configuration Validation — Section III-C — will predict error in session data and could be used to highly tune instruments.

B. Synchronization

Using multiple, independent sensor streams requires determination of a unified virtual time and to synchronize data in the streams on that virtual time. The two streams are recorded at different frequencies and a single HD CMOS generated video frame is not entirely discrete, so perfect synchronization with this equipment is not possible. Synchronization is accomplished by identifying a candidate contact event between WB and enclosure rails and then stepping video and GAZEBO visualization to the same event. For each motion capture session, the following synchronization parameters are recorded: the video frame number of the first collision event that could be synchronized with VICON data, the integrated virtual time of the same collision in the VICON samples, the total number of frames in the video, the frame rate of the video, the sampling rate of VICON, and the frame number where both VICON and video data become stable and usable. Synchronization is then validated by loading the signal data into the GAZEBO visualization with an added LED model to indicate signal events, by computing the virtual time for the frames where a signal event occurred, and then by comparing the simulated visualization of a number of signal events with video to confirm that the state matches in both the visualization and the video.

The first collision event that causes the WB to come to a stop and provides a clear view of the WB orientation is used as the synchronization event. Once a video synchronization event is identified, GAZEBO visualization is stepped to the same event and the video frame number and the VICON state index are recorded. This synchronization event acts as the basis for computing virtual time between the two streams.

To correct for initial sampling error and to compute virtual time, a VICON *starting sample* is selected and virtual time is computed for each VICON sample by incrementing time based on the index of the sample. Because VICON streams data when a registered model is visible to the array, the subject is hidden in the bag until it is inserted into the enclosure; however, there is some error in the first data samples that are generated as the bag may not fully obscure the subject which may lead to erroneous detection. An arbitrary sample is chosen as the starting sample where the VICON samples clearly and consistently align with the video data and samples preceding the starting sample are thrown away to correct this error. This starting sample is used as t_0 in virtual time.

C. Video analysis

The video stream is processed frame by frame to catalog each identifiable recorded signal event. An *identifiable signal* may either be directly observable when the LED is visible to the video camera or be indirectly observable when projected LED light is visible by reflection from the environment. A number of signals are not identifiable due to occlusion of the LED to the camera, so based on the estimated frequency of the motor if a signal is not identifiable but should have occurred a gap is noted. Gaps may consist of the loss of one or more signals. After repeated assessment, we chose not to

TABLE I: HD video LED signal detection

Session	Identifiable Signals	Indirect Signals	Gaps	Two-Frame Signals	Longest Observation (s)
1	249	73	31	39	8.9
2	242	64	32	29	7.567
3	215	37	38	22	4.433
4	232	46	29	18	7.8
5	225	50	34	20	10.33
6	207	35	27	22	8.367
7	265	68	41	46	6.533
8	271	101	37	54	9.567
9	249	58	33	31	6.2
10	265	82	37	39	9.567

For each capture Session, HD video was recorded to capture LED signals generated by the rotating motor assembly. Video was analyzed frame-by-frame for signal events and each signal observation was classified. Identifiable Signal quantifies the number of discrete signal identified on video. Indirect Signals quantifies the number of discrete signal identified through environmental reflection of light. Gaps quantifies the number of times that at least one signal should be observed but no signal was detected. Two-Frame Signals quantifies the number of times a single signal spans consecutive video frames. Longest Observation quantifies the longest period where signals were continuously observed.

attempt to infer the number of signals lost in gaps. Signals may also span two video frames with one of the two signals appearing weak which suggests that the signal duration is approximately equal to the frame rate of the camera plus some small fraction of the frame rate subject to the velocity of the motor assembly — Table I.

V. CONCLUSION

We presented a process for collecting data from a wild, robotic system and provide this data in a repository at `https://www.github.com/PositronicsLab/wild-body` towards simulation validation. The repository includes all raw VICON and video data collected, the SOLIDWORKS models and inertial estimates, GAZEBO controllers developed to validate, synchronize and interpolate, and all signal and synchronization data.

Our motion capture process could be improved in several ways, but evolution of motion capture hardware and processes might naturally diminish error in future data collection. Using a similar approach, more signal data could be collected by either adding video cameras, adding a high speed camera, usage of mirrors, or a combination of optic hardware, and array and marked robot performance could be better validated by integrating Vicon Analysis — Section IV-A — into Configuration Validation — Section III-C.

Future work will concentrate on modeling WB motor assembly physics and on qualitative validation — since telemetry data is essentially guaranteed to differ between simulation and *in situ* — using the collected data. Analytics would be greatly improved through automated synchronization and video analysis using image recognition.

ACKNOWLEDGMENTS

We would like to acknowledge Jack Shannon for assisting in motion capture, Roxana Leontie for creating the circuit, and Jon Torrey for creating the SOLIDWORKS WB model.

REFERENCES

[1] L. Bobadilla, F. Martinez, E. Gobst, K. Gossman, and S. M. Lavalle. Controlling wild mobile robots using virtual gates and discrete transitions. In *American Control Conference(ACC)*, pages 743–749, Montréal, Canada, June 2012.

[2] L. Bobadilla, O. Sanchez, J. Czarnowski, K. Gossman, and S. M. LaValle. Controlling wild bodies using linear temporal logic. In *Proc. Robotics Science and Systems(RSS) VII*, pages 17–24, Los Angeles, USA, June 2011.

[3] A. Boeing and T. Bräunl. Evaluation of real-time physics simulation systems. In *Proceedings of the 5th International Conference on Computer Graphics and Interactive Techniques in Australia and Southeast Asia*, GRAPHITE '07, pages 281–288, New York, NY, USA, 2007. ACM.

[4] D. E. Gierl, L. Bobadilla, O. Sanchez, and S. M. Lavalle. Stochastic modeling, control, and verification of wild bodies. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 549–556, Hong Kong, China, May 2014.

[5] T. Ylikorpi and J. Suomela. Ball-shaped robots. In H. Zhang, editor, *Climbing & Walking Robots, Toward New Applications*, chapter 11, pages 546–567. Itech Education and Publishing, Vienna, Austria, October 2007.