Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A behaviour-centric digital twin framework for animal monitoring: from managed enclosures to landscapes

Basavaraj B. Pujar
Vantara (Greens Zoological Rescue and Rehabilitation Centre)
Keerthikrutha Seetharaman
Vantara (Greens Zoological Rescue and Rehabilitation Centre)
Pinkal Kumar
Vantara (Greens Zoological Rescue and Rehabilitation Centre)
Charan Kumar
Vantara (Greens Zoological Rescue and Rehabilitation Centre)
Ayushman Singh
Vantara (Greens Zoological Rescue and Rehabilitation Centre)
Smitha D Gnanaolivu
Vantara (Greens Zoological Rescue and Rehabilitation Centre)
Brij Kishor Gupta
bkgupta.brij@gmail.com

Vantara (Greens Zoological Rescue and Rehabilitation Centre)

Research Article

Keywords: digital twin, animal behaviour, camera network design, space use, multi-camera fusion, ex situ, wildlife rehabilitation, Unreal Engine

Posted Date: September 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10882389/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background.

Animal monitoring increasingly relies on fixed sensors—CCTV in managed enclosures, camera traps and biologgers in the wild—yet each sensor observes a fragment of the habitat, sensor networks are usually designed ad hoc, and results are analysed per sensor, not in the animal's world. Installation work in occupied enclosures also disturbs animals, so monitoring infrastructure should be designed correctly once. Digital twins are emerging in ecology at the population and ecosystem scale, but none operates at the scale of the individual animal's enclosure or home range.

Methods.

We present a behaviour-centric digital twin framework with a four-stage lifecycle—(i) a *design twin* for one-shot sensor-network design (coverage, occlusion, solar glare, cabling); (ii) a *shadow* that ray-casts multi-camera detections onto the twin geometry into one world frame (validated here with synthetic detections), yielding coverage-normalised three-dimensional space use and movement metrics; (iii) a *twin* closing the decision loop; and (iv) *in-situ transfer* of the same architecture to landscape twins, including continuity of an individual's twin from rehabilitation to release. We instantiate the framework in Unreal Engine 5 for two enclosures at a large wildlife rescue and rehabilitation centre in India (American black bear *Ursus americanus*, five installed cameras; pygmy hippopotamus *Choeropsis liberiensis*) and validate the detection-to-analysis pipeline end-to-end with a seeded synthetic self-test through the virtual cameras.

Results.

For the installed five-camera arrangement, the design twin computes 85.1% of the paddock visible to at least one camera at bear height (57.4% to two or more; 14.9% blind), attributes occlusion to specific structures, and quantifies each camera's contribution, separating area generalists from a deliberate den-door specialist. The self-test (20,302 noisy bounding boxes) recovers ground positions with 0.19 m median error per camera; fusing cameras reduces error to 0.07 m median (0.20 m RMSE) for multi-camera clusters, observes 94.8% of animal-seconds, and reproduces the modified spread-of-participation index within 0.02 of ground truth.

Conclusions.

A geometry-faithful habitat twin converts fragmented per-camera monitoring into world-referenced, coverage-aware ecological measurement, and moves sensor-network design decisions ahead of

installation, where they no longer disturb animals. The framework, aligned with dynamic data-driven digital twin concepts, transfers unchanged from managed enclosures to in-situ landscape monitoring.

1 Background

Automated observation has become the backbone of behaviour and welfare monitoring for animals under managed care and in the wild. Camera-based systems dominate: a recent systematic review of computing technology in zoos and aquariums found camera or video monitoring to be the most frequently used technology (73 instances across the 125 studies surveyed) while reporting no use of digital twins or game-engine-based systems for monitoring [1]. Deep-learning pipelines now localise and identify individual animals [2], track enclosure use [3, 4], and quantify behaviour in one of our focal species [5]; machine learning is reshaping wildlife monitoring generally [6].

Three structural problems, however, limit what this camera-first practice can measure.

First, each camera observes a fragment. In a recent jaguar study, three cameras positioned in collaboration with keeper staff still covered only 30.3% of the enclosure, and space use was necessarily visualised as per-camera image-space heatmaps [4]. Fragmented views conflate “not used” with “not seen”: a zone that no camera covers is indistinguishable, in a per-camera heatmap, from a zone the animals avoid. Multi-camera fusion onto a common ground plane has been demonstrated in livestock barns [7], but with flat two-dimensional homographies, without habitat geometry, and without correcting occupancy for what the network can actually see.

Second, sensor networks are designed ad hoc. Optimal camera placement is a classical problem—from art-gallery visibility [8] through task-specific layout optimisation [9] and coverage modelling surveys [10] to activity-aware placement learned from surveillance data [11]. In animal facilities the only closely related work optimises camera placement for cattle pens using a 3D model and genetic search [12], with static, hand-set importance weights. None of this practice has reached zoos or rescue facilities [1], where placement is negotiated with husbandry staff and fixed by whatever mounting points exist. Critically for animal welfare, installation is not iterative: cabling, drilling and repeated access to an occupied enclosure disturb the animals, so the network should be right the first time. Temporary exhibit design has been proposed exactly because zoos lack means to test designs before committing to them [13].

Third, spatial behaviour metrics remain two-dimensional and coverage-blind. Enclosure-use studies standardly compute the spread of participation index (SPI) over zones, in its modified form weighting zones by area [14, 15], or resource electivity [16]; shade availability mapping added a thermal dimension [17]. Three-dimensional kernel methods exist for wildlife telemetry [18], and 3D photogrammetry has recently been used to quantify usable space in primate habitats—explicitly without mapping animal positions into the model, which its authors name as the needed next step [19]. That step—animal positions measured *in* a faithful 3D habitat model, corrected for what the sensors can see—has not been taken.

Digital twins offer the missing substrate. Originating in industrial cyber-physical systems, twins are entering ecology rapidly: the TwinEco framework defines ecological twins as dynamic data-driven systems coupling observation, understanding and action [20]; FAIR-oriented biodiversity twins [21] and a national-scale twin forecasting bird biodiversity from citizen-science data [22] demonstrate the concept at the population scale; the trend has reached general scientific press [23]. In parallel, precision livestock farming builds twins of barns and animals [24, 25, 26, 27], most recently coupling video behaviour recognition to a dairy twin [28], and conservation practice has used population simulations labelled as twins to design camera-trap deployments against animal–vehicle collisions [29]. Game engines specifically are already established in animal computer vision as generators of annotated synthetic imagery [30, 31], but not as operational monitoring infrastructure. Between the ecosystem-scale twins and the barn twins lies an unoccupied scale: the *habitat of the individual animal*—a zoo or rehabilitation enclosure, or a home range—represented with faithful 3D geometry and used as the reference frame for monitoring itself.

This paper fills that gap with four contributions:

1. a **lifecycle framework** (design twin → shadow → twin → in-situ transfer) that places a geometry-faithful, interactive habitat twin at the centre of animal monitoring, aligned with the TwinEco/DDDAS view of ecological twins [20], with an explicit maturity ladder that states which components are model, shadow, or twin at any time;
2. a **design-stage methodology** that decides camera count, placement, and cabling before installation: occlusion-tested, animal-height coverage with per-structure occlusion attribution; detectability limits; camera-failure redundancy; critical-region visibility; and—to our knowledge for the first time in animal monitoring—per-camera *solar glare timelines* computed from the site's real sun geometry;
3. a **shadow-stage pipeline** that ray-casts per-camera detections onto the twin geometry into one world frame, fuses cameras with uncertainty weighting and plausibility filters, and produces *coverage-normalised* occupancy plus standard enclosure-use indices (modified SPI, electivity) on 3D resource zones—so that unobservable space is reported as unobservable rather than unused;
4. an **implementation** for two real enclosures of a large rescue and rehabilitation centre, with quantitative **validation** on the black bear paddock: a seeded, fully reproducible synthetic self-test quantifies end-to-end position and index recovery, establishing the accuracy envelope of the pipeline ahead of operational footage.

2 Methods

2.1 Sites and twin construction

The framework is implemented for two enclosures at Vantara (Greens Zoological Rescue and Rehabilitation Centre), Jamnagar, India (22.326°N, 69.923°E): a 990.28 m² American black bear (*Ursus americanus*) paddock (“Paddock 23”), home to two co-released rescued individuals—Ballu, a black-morph male, and Coco, a cinnamon-morph individual (Fig. 2c)— and a 1,075.54 m² pygmy hippopotamus

(*Choeropsis liberiensis*) paddock. Both enclosures are actively monitored subjects: the black bear paddock carries an operational on-device monitoring pipeline analysed in a companion study [32], and automated tracking has recently been applied to captive pygmy hippopotamuses [5]. Enclosure geometry was derived from satellite and drone imagery plus ground photographs (Fig. 2a,b): boundary polygons extracted from geo-referenced imagery; heights, structures, pond bathymetry and night-den layout reconstructed from survey photographs; Fig. 2d,e show the corresponding areas in the twin. Each enclosure is a level of one Unreal Engine 5.8 (Epic Games) application; the Paddock 23 level comprises the terrain with trails and a 1.0 m-deep pond, a 6.75 m rock wall with five night-den doors, a stacked-slab rock cascade, roof structure, perimeter fencing, vegetation, climbing and enrichment structures, and the five camera positions of the site's CCTV network—five fixed RTSP cameras (1920×1088 at 25 fps; 1088 is the encoder's 16-pixel-aligned frame height) pole-mounted at 3.4–4.6 m [32]—as virtual cameras, positioned and oriented per the installed arrangement documented by the installation team (a corner camera beside the night house, a pole-mounted den-door camera inside the paddock, two boundary cameras at the pool and gallery ends, and a visitors-gallery camera; represented at a uniform 3.66 m mounting height and 110.4° horizontal field of view pending metric as-built calibration against the twin using ground control points on the enclosure structures). A directional sun driven by the site latitude, longitude and local time reproduces real solar geometry (validated against expected solar elevation, e.g. 84.9° at 12:30 local clock time on 21 June, about 20 min before local solar noon). Physical fidelity is treated as a gate: enclosure areas are carried from the geo-referenced extraction, and a similarity registration of the twin's boundary against the north-up geo-referenced outline (tested over both chiralities) shows a 0.50 m mean residual and confirms the twin's compass orientation to within 0.5°; a height/structure dimension audit and registered virtual-versus-real camera-view overlays complete the acceptance checklist at operational deployment (Sect. 2.2.1). Quantitative results in this paper are reported for Paddock 23, whose camera network is installed; the hippopotamus twin implements the same geometry, camera-rig, solar and analytics stack and enters the identical analyses once its camera plan is finalised.

2.2 Framework

Figure 1a shows the lifecycle. **Stage 1, design twin**: before any hardware exists (or when a network is re-planned), the twin answers where cameras should go—under constraints no 2D plan captures: occlusion by three-dimensional structures and vegetation, animal size and camera resolution, seasonal sun. Because installation work in an occupied enclosure disturbs animals, iteration happens virtually and installation happens once. **Stage 2, shadow**: once cameras stream, their detections are projected onto the twin's geometry into a single world frame (Sect. 2.4); the twin mirrors the physical system without yet acting on it. **Stage 3, twin**: analytics feed decisions—alerts (e.g. prolonged submersion, animal not detected), seeded what-if simulation (an enrichment moved, a camera offline), replay of recorded sessions—with actuation via husbandry staff (human-in-the-loop). **Stage 4, in-situ transfer**: the identical architecture runs on landscape twins built from elevation models and seasonal environmental layers, ingesting GPS collars, camera traps, or acoustic sensors instead of CCTV; for a rescue and rehabilitation

institution this enables *continuity of the individual's twin* from the pre-release enclosure (behavioural baselines under full observation) to the post-release landscape (sparse telemetry), a capability we design for and discuss in Sect. 4.1.

Figure 1b shows the architecture. A data hub owns a twin clock and typed, versioned message streams (EntityPose, SensorSample, CameraHealth, Event); data sources are pluggable—*Simulated* (a seeded behaviour model), *Replay* (recorded sessions), or *Live* (detections, telemetry, sensors)—behind the same schema, so the same build moves from simulation to operations by configuration, not redesign. Every session is recorded append-only with the seed and parameter hash in the header; identical seeds reproduce identical runs, and analytics computed from a replay must equal the live run (determinism and replay-integrity tests). This follows the dynamic data-driven (DDDAS) structure that TwinEco identifies as the core of ecological twins [20].

2.2.1 Maturity ladder

Digital-twin claims in ecology are frequently ahead of their implementations; we therefore state maturity per component (Table 1) using the conventional model–shadow–twin distinction: a *digital model* has no automatic data flow; a *digital shadow* updates automatically from the physical system; a *digital twin* additionally feeds decisions back. Under this taxonomy, the present system is a validated digital model with shadow-ready interfaces: the geometry, camera rig, solar model, analytics and record/replay layers are operational; live ingestion of the installed cameras' detection stream is the step in progress (Sect. 4.2).

Table 1
Maturity of framework components at the time of writing (model / shadow / twin; evidence in this paper unless noted).

p{0.30\textwidth}p{0.13\textwidth}p{0.48\textwidth}@{} Component	Status	Evidence
Enclosure geometry (2 sites)	model	geo-referenced extraction; boundary registration 0.50 m mean residual, orientation 0.5°; photo-vs-twin comparison (Fig. 2)
Virtual camera rig (5 cams, Paddock 23)	model	installed arrangement (team site sketch); metric as-built survey pending
Solar model	model	site solar geometry; paddock orientation verified against north-up imagery (0.5°)
Coverage/glare analyser	operational	Sect. 3.1–3.2, Figs. 3–5
Detection→world projection + fusion	validated	synthetic self-test, Sect. 2.5, Figs. 6–8
Space-use indices (SPI, electivity) on 3D zones	validated	index recovery within 0.02, Sect. 3.4
Record / replay / determinism	operational	seeded runs reproduce byte-identical logs
Detection pipeline on the installed network	operational	on-device YOLOv8 + ViT re-identification (99.2% inference-system uptime); companion study [32]
Projection of that stream into the twin (shadow)	in progress	calibration + ray-cast of the companion pipeline's segments
Alerts / what-if decision loop (twin)	designed	Sect. 2.2; human-in-the-loop
Landscape twin (in-situ)	designed	Sect. 4.1

2.3 Design-stage analyses

All analyses operate on a 0.5 m grid over the enclosure boundary polygon. For each cell, the walkable surface height is found by a vertical ray-cast against the twin geometry (cells on top of structures are excluded); the visibility target is placed at shoulder height of the focal species (0.7 m, within the adult

American black bear range). For each camera and cell we evaluate three nested predicates: *in field of view* (pinhole projection inside the image); *detectable* (a 1.5 m animal subtends ≥ 40 px horizontally and range ≤ 45 m—thresholds configurable to the detector in use); and *visible* (a ray from the camera to the target is not blocked). Ray-casts run against the full scene geometry with one ecological refinement: ground-cover vegetation (grass and herb layers) is treated as transparent—a bear is not hidden by 0.5 m grass tufts that intersect a sight-line metres in front of the lens—while trees, rock, logs and structures block. Each blocked ray records the *material* of its first blocker, yielding an occlusion attribution histogram that tells designers what to move or accept (Sect. 3.1). Aggregates include per-camera and union coverage, k-camera redundancy, leave-one-camera-out coverage (robustness to failure), per-zone coverage, and a critical-region visibility matrix over husbandry-relevant targets (pond, den doors, shade, enrichment structures, rock cascade).

Solar glare. For four nominal equinox and solstice dates (21 March, June, September, December) at 30 min steps, the sun vector from the site's solar geometry is classified per camera as *sun-in-frame* (sun disc projects inside the image), *backlit* (the three-dimensional angle between the optical axis and the sun direction is below 0.75× the horizontal field of view, i.e. 82.8°, while the sun disc is outside the frame), *side-lit*, or *front-lit*. Cameras facing the equator at this latitude (22°N) are structurally exposed (the paddock's compass orientation is verified against north-up imagery to 0.5°, Sect. 2.1); the predicted timeline indicates when each camera's footage is expected to be degraded and lets mounting decisions (or infrared bias) be taken before installation.

2.4 Shadow-stage pipeline: detections to world-referenced space use

The pipeline consumes per-camera, per-frame bounding boxes (camera, time, box, track id, confidence)—the standard output of animal detectors [2, 4], and exactly the record produced by the enclosure's operational edge-AI pipeline, whose verified segments carry camera, tracker identifier, timestamps and individual identity [32]—and needs only each camera's pose and intrinsics, both known in the twin. For each detection, the ray through the box's bottom-centre (the foot point; or the box centre with a centroid-height correction, when only a centre point is exported) is cast against the twin geometry; the first blocking hit is the animal's ground position, in world coordinates, including on sloped terrain, in the pond shallows, or on climbing structures—cases where flat-homography mapping [7] breaks. Each position carries an uncertainty proxy: the ground footprint of one pixel at that range and depression angle. Two plausibility filters, both available for real footage, reject degenerate rays: hits outside the enclosure boundary, and hits whose ray distance disagrees by more than 40% with the distance implied by the box height (given the species' shoulder height). Positions from different cameras are then fused in 1 s bins by proximity clustering (1 m radius) with inverse-variance weighting and a robust pruning step that removes members inconsistent with the cluster centroid (guarding against merging two different animals).

Coverage-normalised occupancy. Fused positions accumulate on the same 0.5 m grid as the coverage analysis. Raw occupancy is divided by each cell's camera visibility count from the design-stage analysis, and cells visible to no camera are flagged *unobservable* rather than reported as zero—making the “not seen vs. not used” distinction explicit in every map. Zone statistics follow the conventions of the companion space-use analysis [32]: per-individual dwell time (minutes and % of detected time), the spread of participation index—in its modified, area-weighted form [14]—Shannon diversity H [33] with Pielou's evenness J [34], electivity [16], between-individual overlap via the Bhattacharyya coefficient [35], and co-detection counts (both individuals in one zone within 60 s). All are computed over 3D resource zones defined on the twin (pond deep/shelf/rim, shade, den front, individual enrichment structures, tree shade, open ground), with zone availability measured as area within the boundary.

2.5 Validation: seeded synthetic self-test

Because the ray-casting, filtering, fusion and index pipeline will be trusted with welfare-relevant conclusions, we validate it end-to-end against known ground truth before operational data. A seeded behaviour model generates 1 Hz trajectories for two agents over one hour (dwell bouts at resource hotspots—pond, shade, platforms, den front—connected by walks at 0.6 m s^{-1}, all inside the boundary). Each ground-truth position is rendered into every camera through the same projective geometry the real cameras have: if the animal is in frame and its sight-line unoccluded, a bounding box is emitted with Gaussian pixel noise ($\sigma = 2$ px). The resulting 20,302 noisy detections are processed by the *identical* inverse pipeline used for real footage, and recovered positions, occupancy maps, and indices are compared with ground truth. The test is deterministic (fixed seeds) and regenerates exactly; it measures the pipeline's geometric error envelope, not animal behaviour.

3 Results

3.1 Design twin: coverage, occlusion attribution, redundancy

Figure 2 shows the enclosure and its twin; Fig. 3 the coverage field of the installed five-camera arrangement (metric as-built calibration pending, Sect. 2.1) over 3,702 analysable cells (925.5 m^2; structure tops excluded). At bear height with occlusion tested, 85.1% of the paddock is visible to at least one camera, 57.4% to two or more, and 30.4% to three or more; 14.9% is blind, mainly along the den front and the paddock margins. The arrangement mixes roles, and the twin quantifies the mix (Table 2): the pole-mounted den-door camera C1 is a deliberate specialist—only 12.6% of the paddock lies in its field of view at all, and it contributes 9.3% area coverage, but with it in place the den-door region reaches 100% dual coverage (C1 itself sees 80% of the door sample points; C3–C5 contribute 60–80%); the boundary and gallery cameras are generalists spanning 34.9–50.4%. Leave-one-out analysis identifies C3 (pool end) as the critical unit—union coverage falls to 69.7% without it, and 15.4% of the paddock is visible to C3 alone—while removing C4 costs only 1.0 percentage point (84.1% remaining), making it the least

disruptive camera to re-site if dual coverage is wanted elsewhere. Because the network was installed before the twin analysis, these design-stage results constitute a retrospective audit of the as-built arrangement; the audit's recommendations (the C4 re-siting case, the infrared priorities of Sect. 3.2) are prospective, and the hippopotamus enclosure—whose camera plan is not yet installed—is the framework's prospective one-shot design case. Occlusion attribution identifies trees as the dominant blocker (canopy 1,537 and trunk 324 blocked cell–camera pairs; canopies modelled as solid, hence conservative), followed by the gunite rock wall (547) and the log structures (256), of 2,888 blocked cell–camera pairs in total (551 of the 3,702 cells are blind to all cameras); ground vegetation, correctly, does not occlude. Coverage is insensitive to the uniform-height assumption: recomputed end-to-end at the extremes of the documented 3.35–4.57 m mounting range, union coverage stays within 85.1–85.8% and dual coverage within 55.0–58.4%. The critical-region matrix (Fig. 5) verifies 100% single-camera visibility of all eight husbandry-relevant regions—including the den doors at 100% network dual coverage, the purpose C1 exists to serve—with the climbing frame the weakest dual-covered resource (25%).

Table 2
Per-camera design-stage metrics for the Paddock 23 five-camera configuration (0.5 m grid, animal height 0.7 m, detectability ≥ 40 px for a 1.5 m animal, ≤ 45 m).

lcccc@{}} Camera	In FOV (%)	Detectable (%)	Visible (%)	Union coverage without it (%)
C1 (night-house gate, pole)	12.6	12.6	9.3	83.4
C2 (paddock corner)	85.1	64.6	37.7	80.8
C3 (pool / pond end)	92.5	56.9	47.6	69.7
C4 (gallery / shade end)	90.0	56.3	34.9	84.1
C5 (visitors gallery)	73.6	67.5	50.4	79.8
Union	85.1% visible ≥ 1 camera; 57.4% ≥2; 30.4% ≥3; 14.9% blind			

3.2 Design twin: solar glare

At 22°N the sun occupies the southern sky most of the year, so the two cameras that face the night-house (south) side are structurally exposed. These are geometric predictions from the twin's solar model, not measurements on footage (Fig. 4): the model places the den-door camera C1 in backlit or sun-in-frame conditions for 6.5 h per day at the equinoxes and 10.0 h at the winter solstice, and the visitors-gallery camera C5 for 6.0 h and 7.5 h respectively; the corner and boundary cameras (C2–C4)

remain front- or side-lit through most operational hours. Such predicted timelines convert a commonly discovered post-installation defect—washed-out footage during exactly the morning activity peak—into a pre-installation design variable (mount re-orientation, shading, or infrared-priority scheduling); verifying the predictions against the installed cameras' footage is part of the planned as-built validation.

3.3 Pipeline validation: position recovery

Of 20,302 synthetic detections, 4.3% produced no geometry hit and the plausibility filters rejected a further 8.5% (0.2% outside the boundary; 8.3% box-height/ray-distance mismatch, dominated by near-horizon foot rays), leaving 17,702 positions (87.2%). Single-camera ground positions are accurate to 0.19 m median (RMSE 1.28 m, p90 1.10 m; the RMSE exceeds the 90th percentile because it is dominated by the sparse long-range tail beyond p90), with error growing steeply beyond ≈ 25 m of range (Fig. 8)—quantifying, for this enclosure size and camera resolution, where single-camera reading of position becomes unreliable. Cross-camera fusion resolves it: cluster estimates informed by two or more cameras—61.7% of the 8,720 fused position estimates, containing 81.2% of the underlying detections—recover positions at 0.07 m median error (RMSE 0.20 m). Including single-member clusters, the overall fused median remains 0.11 m while the RMSE rises to 1.71 m. This is a population-reweighting effect rather than fusion damage, and the two RMSEs are not directly comparable: the 3,336 detections that no second camera corroborated (disproportionately long-range; RMSE 2.76 m) each count once among 8,720 cluster estimates instead of once among 17,702 detections, roughly doubling the weight of the same long-range tail. Cross-identity merges—clusters mixing the two agents within the 1 m radius—are rare and benign after robust pruning: 236 clusters (2.7%), RMSE 0.23 m against 0.20 m for pure multi-camera clusters. Both conditions are flagged by the cluster metadata (member count, dispersion) that accompanies every fused position. Across the hour, 94.8% of ground-truth animal-seconds were observed by at least one camera—consistent with the design-stage finding that the frequently used resource zones are well covered (overall area coverage is 85.1%; animals favour covered zones).

3.4 Space use: from per-camera fragments to coverage-normalised maps and indices

Figure 6 shows the same hour as (i) ground truth, (ii) single-camera fragments, (iii) the fused world-frame map, and (iv) the coverage-normalised map with unobservable cells rendered as such. Single-camera maps miss entire resource zones (the den-door camera C1 alone records little beyond the den approach; the corner camera C2 misses the pond end); the fused map reproduces the ground-truth structure. Zone-level measures follow (Fig. 7): the modified SPI computed from fused positions is 0.623 against 0.640 for ground truth (single-camera values deviate further and differ between cameras), and the electivity ranking of zones (the enrichment structures and the deep pond preferred; open ground and the den front avoided) is preserved in the fused estimate—identical most- and least-preferred zones, with a single adjacent-rank swap mid-table. The twin thus recovers the indices that zoo-biology practice

already uses [15], but measured in world coordinates, corrected for coverage, and defined on 3D resource zones.

4 Discussion

What the twin changes

Relative to camera-first practice, the framework moves three decisions to firmer ground. *Design happens before disturbance*: the coverage, redundancy, attribution and glare analyses of Sect. 3.1–3.2 were produced entirely in the twin—for Paddock 23 as a retrospective audit of an installed network whose recommendations are prospective—and each has a direct design consequence (C4 is the cheapest camera to re-site for dual coverage of the climbing frame; the south-facing den-door and gallery cameras are the infrared-priority candidates; canopy occlusion is quantified and accepted as ecologically inherent; the den-door specialist's area-versus-task trade-off is made explicit rather than discovered). This operationalises, for monitoring infrastructure, the “test before you build” agenda argued for exhibit design [13], with the difference that the tests here are quantitative and reproducible. *Measurement happens in the animal's frame*: foot-point ray-casting onto real geometry generalises ground-plane homography fusion [7] to sloped, stepped and submerged surfaces, and directly supplies the 3D space-use data that habitat-photogrammetry work identified as its missing complement [19]. *Reported use is conditioned on observability*: coverage normalisation prevents the silent equation of unobserved with unused—a bias present in any per-camera heatmap, including the 30%-coverage example that motivated this work [4]. The companion analysis of this same enclosure makes the need concrete: working per camera sector on manually verified detections, it reports strong pool-sector concentration for both individuals (58.1% and 35.2% of detected time) with extensive overlap (Bhattacharyya coefficient 0.93), but must assume equal-area spatial units for its spread-of-participation index and lacks a per-camera recording/observability record from which to form an occupancy denominator (the 99.2% figure it reports describes inference-system uptime, not camera observation time) [32]. The twin supplies precisely these missing terms—measured zone areas for the modified SPI (Sect. 2.4) and per-cell visibility for the denominator (Sect. 2.3)—and replaces the sector-per-camera partition with one world frame.

The self-test quantifies an accuracy envelope sufficient for the intended analyses: 0.07–0.19 m median position error is far below the body length of either focal species and below the 0.5 m analysis grid, so dwell, zone occupancy, movement and proximity metrics are limited by behaviour sampling, not geometry. The envelope also delimits what single cameras can and cannot support in an enclosure of this size (reliable position to ≈ 25 m; beyond that, fusion or closer mounts).

Relation to ecological digital twins

The framework is deliberately an instance of the TwinEco conception—modular components in observation–understanding–action layers, dynamic model–data coupling, explicit state and feedback [20]—applied at a scale that ecological twin initiatives have not yet occupied [21, 22]. It complements,

rather than competes with, population- and ecosystem-scale twins: the enclosure twin measures individuals at centimetre scale under near-complete observation, the regime in which behavioural ground truth can be established before methods transfer to the sparse-observation regime of the wild. Conversely, it imports from livestock twins [25, 28] the discipline of production deployment (deterministic replay, schema-stable interfaces), which zoo and conservation software mostly lacks.

4.1 In-situ transfer and rescue–release continuity

Stage 4 is a design in this paper, not a result, and we state it as such. The architecture transfers without structural change: a landscape twin substitutes elevation and seasonal layers (vegetation, water, temperature) for surveyed enclosure geometry; GPS collars, camera traps and acoustic arrays replace CCTV behind the same message schema; the coverage engine answers sensor-placement questions of the kind currently addressed with population simulations [29]; and 3D space-use estimation connects to established movement-ecology methods [18] and biologging practice [36, 37]. For a rescue and rehabilitation institution the scientifically distinctive opportunity is *continuity*: the same individual, one identity, observed first in an enclosure twin under near-total coverage (behavioural baseline, welfare indices) and then in a landscape twin after release (movement, habitat selection, survival correlates). We know of no published monitoring system that maintains an individual's digital twin across that boundary; the black bear paddock studied here—housing rescued, co-released individuals [32]—is the intended first case.

4.2 Limitations

Four limitations bound the present claims. (1) The behavioural trajectories used for validation are synthetic; they establish geometric and statistical correctness of the pipeline, not behavioural findings. The enclosure's operational pipeline already produces manually verified real detections (932 segments, 10.4 h over 23 recording days; [32]), and its raw per-frame exports carry image coordinates; projecting a verified detection stream through the calibrated twin—and re-estimating Figs. 6–7 on it—is the immediate next step and will accompany a revised version of this work. (2) Tree canopies are modelled as solid occluders, making coverage conservative for a partially translucent canopy; occlusion attribution makes this assumption explicit and auditable. (3) Fidelity is currently anchored in geo-referenced geometry with a registered boundary (0.50 m mean residual, orientation 0.5°), solar validation, and qualitative photograph-versus-twin comparisons (Fig. 2); the height/structure dimension audit and registered view overlays complete with the operational deployment. (4) Results cover two enclosures at one institution and one camera technology; the methodology is species- and sensor-agnostic by construction, but generality is asserted for the method, not yet demonstrated across facilities.

5 Conclusions

A geometrically faithful, interactive habitat twin turns animal monitoring from a collection of camera views into a measuring instrument: sensor networks are designed once, virtually, under occlusion, redundancy and solar constraints; detections from many partial views become world-referenced

positions with quantified uncertainty; and space-use maps and welfare indices become coverage-aware. Validated here end-to-end against reproducible synthetic ground truth on an American black bear enclosure twin—and built on interfaces that accept live data by configuration—the framework is positioned to carry the same measurement discipline from managed enclosures to in-situ landscapes, including the unbroken monitoring of individuals from rehabilitation to release.

Declarations

Ethics approval and consent to participate: This study involved no procedures on animals: all results derive from simulation within the digital twin and from infrastructure planning; no animal data were collected for this manuscript. [TBD: institutional statement; update when operational footage is added.]

Consent for publication: Not applicable.

Availability of data and materials: The analysis code (coverage/glare analyser, projection and fusion pipeline, self-test, figure generation), configuration files, and the numeric outputs underlying all figures and tables (and per-camera coverage maps, provided as supplementary material) are available from the corresponding author and will be deposited in a public repository upon acceptance [TBD: repository DOI]. The synthetic validation data regenerate deterministically from the released code and seeds.

Competing interests: The authors declare no competing interests.

Funding: [TBD].

Authors' contributions: [TBD].

Acknowledgements

We thank the animal care and veterinary teams at Vantara for husbandry knowledge that shaped the critical-region definitions, and the Vantara team for the site and animal photographs in Fig. 2. We than Mr. Vivaan Karani, Chief Executive Officer, Vantara and Mr. Aakash Pratap Singh, Chief Operating Officer, Vantara for constant encouragement and all logistical help provided during the present study.

References

1. Hassinger L, Nippert-Eng C (2025) Beyond the camera trap: a systematic review of computing technology used to monitor and interact with (more) varied taxa in zoos and aquariums. Animals 15(12):1721. https://doi.org/10.3390/ani15121721
2. Zuerl M, Stoll P, Brehm I, Sueskind J, Raab R, Petermann J, Zanca D, Simon R, von Fersen L, Eskofier B (2024) Automated long-term monitoring of stereotypical movement in polar bears under human care using machine learning. Ecol Inf 83:102840. https://doi.org/10.1016/j.ecoinf.2024.102840

3. Gabaldon J, Zhang D, Lauderdale L, Miller L, Johnson-Roberson M, Barton K, Shorter KA (2022) Computer-vision object tracking for monitoring bottlenose dolphin habitat use and kinematics. PLoS ONE 17:e0254323. https://doi.org/10.1371/journal.pone.0254323
4. Larsen LLN, Christensen N, Kristensen T, Faddersbøll TL, Lassen ARW, Rasmussen B, Pagh S, Pertoldi C (2026) Automated activity tracking and space use monitoring of captive jaguars with machine learning. Animals 16(10):1504. https://doi.org/10.3390/ani16101504
5. Nielsen J, Gammelgård F, Lund SM, Præstekær ASB, Vinterberg Frandsen A, Strandqvist C, Nielsen MH, Olsen RN, Pagh S, Pertoldi C (2026) Olfactory enrichment of captive pygmy hippopotamuses with applied machine learning. Animals 16(3):385. https://doi.org/10.3390/ani16030385
6. Tuia D, Kellenberger B, Beery S, Costelloe BR, Zuffi S, Risse B, Mathis A, Mathis MW, van Langevelde F, Burghardt T, Kays R, Klinck H et al (2022) Perspectives in machine learning for wildlife conservation. Nat Commun 13:792. https://doi.org/10.1038/s41467-022-27980-y
7. Nasir MF, Fuentes A, Han S, Liu J, Jeong Y, Yoon S, Park DS (2025) Multi-camera fusion and bird-eye view location mapping for deep learning-based cattle behavior monitoring. Artif Intell Agric 15:724–743. https://doi.org/10.1016/j.aiia.2025.06.001
8. Chvátal V (1975) A combinatorial theorem in plane geometry. J Comb Theory Ser B 18:39–41. https://doi.org/10.1016/0095-8956(75)90061-1
9. Erdem UM, Sclaroff S (2006) Automated camera layout to satisfy task-specific and floor plan-specific coverage requirements. Comput Vis Image Underst 103:156–169. https://doi.org/10.1016/j.cviu.2006.06.005
10. Mavrinac A, Chen X (2013) Modeling coverage in camera networks: a survey. Int J Comput Vision 101:205–226. https://doi.org/10.1007/s11263-012-0587-7
11. Wu H, Zeng Q, Guo C, Zhao T, Chen CW (2024) Target-aware camera placement for large-scale video surveillance. IEEE Trans Circuits Syst Video Technol 34:13338–13348. https://doi.org/10.1109/TCSVT.2024.3445151
12. Sourav AA, Peschel JM (2022) Visual sensor placement optimization with 3D animation for cattle health monitoring in a confined operation. Animals 12(9):1181. https://doi.org/10.3390/ani12091181
13. Brereton JE, Coe J, Fernandez EJ (2025) Improving zoo exhibit design: why we need temporary exhibit design. J Zoological Bot Gardens 6(1):19. https://doi.org/10.3390/jzbg6010019
14. Plowman AB (2003) A note on a modification of the spread of participation index allowing for unequal zones. Appl Anim Behav Sci 83:331–336. https://doi.org/10.1016/S0168-1591(03)00142-4
15. Brereton J, Fernandez E (2022) Which index should I use? A comparison of indices for enclosure use studies. Anim Behav Cognition 9(1):119–132. https://doi.org/10.26451/abc.09.01.10.2022
16. Vanderploeg HA, Scavia D (1979) Two electivity indices for feeding with special reference to zooplankton grazing. J Fish Res Board Can 36:362–365. https://doi.org/10.1139/f79-055
17. Wark JD, Wierzal NK, Cronin KA (2020) Mapping shade availability and use in zoo environments: a tool for evaluating thermal comfort. Animals 10(7):1189. https://doi.org/10.3390/ani10071189

18. Tracey JA, Sheppard J, Zhu J, Wei F, Swaisgood RR, Fisher RN (2014) Movement-based estimation and visualization of space use in 3D for wildlife ecology and conservation. PLoS ONE 9(7):e101205. https://doi.org/10.1371/journal.pone.0101205
19. Ritzler CP, Mather SV, Johnson IT, Dennis PM, Lukas KE, Koester DC (2025) Using 3D photogrammetry to quantify usable space in zoo-housed primate habitats. Am J Primatol 87:e70002. https://doi.org/10.1002/ajp.70002
20. Khan T, de Koning K, Endresen D, Chala D, Kusch E (2025) TwinEco: A unified framework for dynamic data-driven digital twins in ecology. Ecol Inf 91:103407. https://doi.org/10.1016/j.ecoinf.2025.103407
21. Islam S, Koivula H, Andrew C, Lopez Gordillo J, Weiland C, Schigel D, Endresen D, Arvanitidis C, Chadwick E, Soiland-Reyes S et al (2026) FAIR digital twins for biodiversity: enabling data, model, and workflow integration. https://doi.org/10.1038/s44185-025-00116-3. npj Biodiversity 5
22. Ovaskainen O, Winter S, Tikhonov G, Lauha P, Lehtiö A, Nokelainen O, Abrego N et al (2026) A digital twin for real-time biodiversity forecasting with citizen science data. Nat Ecol Evol 10:481–495. https://doi.org/10.1038/s41559-025-02966-3
23. Richardson H (2025) Digital flocks, predators allow ecologists to simulate real-world animal behaviour. Nature 646:763–765. https://doi.org/10.1038/d41586-025-03314-y
24. Neethirajan S, Kemp B (2021) Digital twins in livestock farming. Animals 11(4):1008. https://doi.org/10.3390/ani11041008
25. Jeong DY, Jo SK, Lee IB, Shin H, Kim JG (2023) Digital twin application: making a virtual pig house toward digital livestock farming. IEEE Access 11:121592–121602. https://doi.org/10.1109/ACCESS.2023.3313618
26. Han X, Lin Z, Clark C, Vucetic B, Lomax S (2022) AI based digital twin model for cattle caring. Sensors 22(19):7118. https://doi.org/10.3390/s22197118
27. Youssef A, Vodorezova K, Aarts Y, Agbeti WEK, Palstra AP, Foekema E, Aguilar L, da Silva Torres R, Grübel J (2024) IUMENTA: A generic framework for animal digital twins within the Open Digital Twin Platform. https://doi.org/10.48550/arXiv.2411.10466. arXiv preprint arXiv:2411.10466
28. Rao S, Garcia E, Neethirajan S (2026) Video-based cattle behaviour detection for digital twin development in precision dairy systems. npj Veterinary Sci 1. https://doi.org/10.1038/s44433-026-00004-x
29. Moulherat S, Pautrel L, Debat G, Etienne MP, Gendron L, Hautière N, Tarel JP, Testud G, Gimenez O (2024) Biodiversity monitoring with intelligent sensors: an integrated pipeline for mitigating animal-vehicle collisions. Nat Conserv 57:103–124. https://doi.org/10.3897/natureconservation.57.108950
30. Plum F, Bulla R, Beck HK, Imirzian N, Labonte D (2023) replicAnt: a pipeline for generating annotated images of animals in complex environments using Unreal Engine. Nat Commun 14:7195. https://doi.org/10.1038/s41467-023-42898-9
31. Tobin J, Fong R, Ray A, Schneider J, Zaremba W, Abbeel P (2017) Domain randomization for transferring deep neural networks from simulation to the real world. In: 2017 IEEE/RSJ International

Conference on Intelligent Robots and Systems (IROS), pp 23–30. https://doi.org/10.1109/IROS.2017.8202133

32. Singh A et al (2026) Continuous individual-level monitoring of rescued American black bears with an on-device edge-AI pipeline: verified enclosure space use of two co-released individuals. arXiv
33. Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27:379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x
34. Pielou EC (1966) The measurement of diversity in different types of biological collections. J Theor Biol 13:131–144. https://doi.org/10.1016/0022-5193(66)90013-0
35. Bhattacharyya A (1943) On a measure of divergence between two statistical populations defined by their probability distributions. Bull Calcutta Math Soc 35:99–109
36. Rudd JL, Collins MA, Righton D, White HL, Witt MJ, Wright S, Hawkes LA (2026) Six decades of animal accelerometry: trends, applications, and future directions. https://doi.org/10.1186/s40462-026-00641-1. Movement Ecology 14
37. Aaser MF, Staahltoft SK, Andersen M, Alstrup AKO, Sonne C, Bruhn D, Frikke J, Pertoldi C (2024) Using activity measures and GNSS data from a virtual fencing system to assess habitat preference and habitat use of dairy cattle. Animals 14(10):1506. https://doi.org/10.3390/ani14101506

Figures

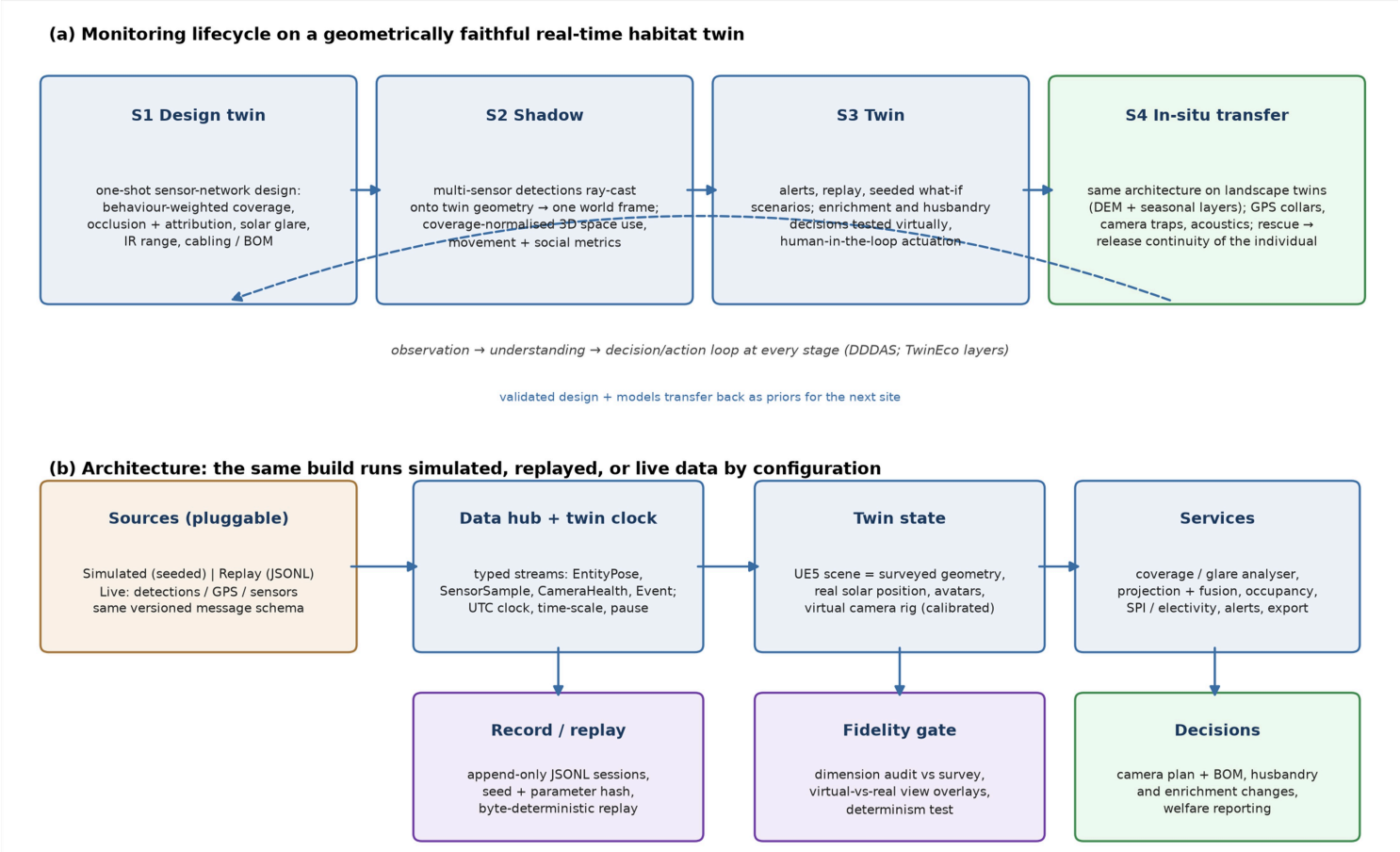


Figure 1

The framework. (a) Monitoring lifecycle on a geometry-faithful habitat twin: design twin (one-shot sensor-network design), shadow (multi-sensor fusion into one world frame), twin (decision loop), and in-situ transfer, with validated designs and models returning as priors for subsequent sites. (b) Architecture: pluggable data sources behind one versioned schema; a data hub with twin clock; the Unreal Engine twin state; analytic services; deterministic record/replay; and a fidelity gate.



Figure 2

The real Paddock 23 and its digital twin. (a) The pond area and (b) the night-house wall photographed on site; (d, e) the corresponding areas in the twin (Unreal Engine 5.8); (c) the resident co-released American black bears, Ballu (black morph) and Coco (cinnamon morph); (f) twin overview with roof structure and perimeter fence. Photographs: Vantara animal-care team.

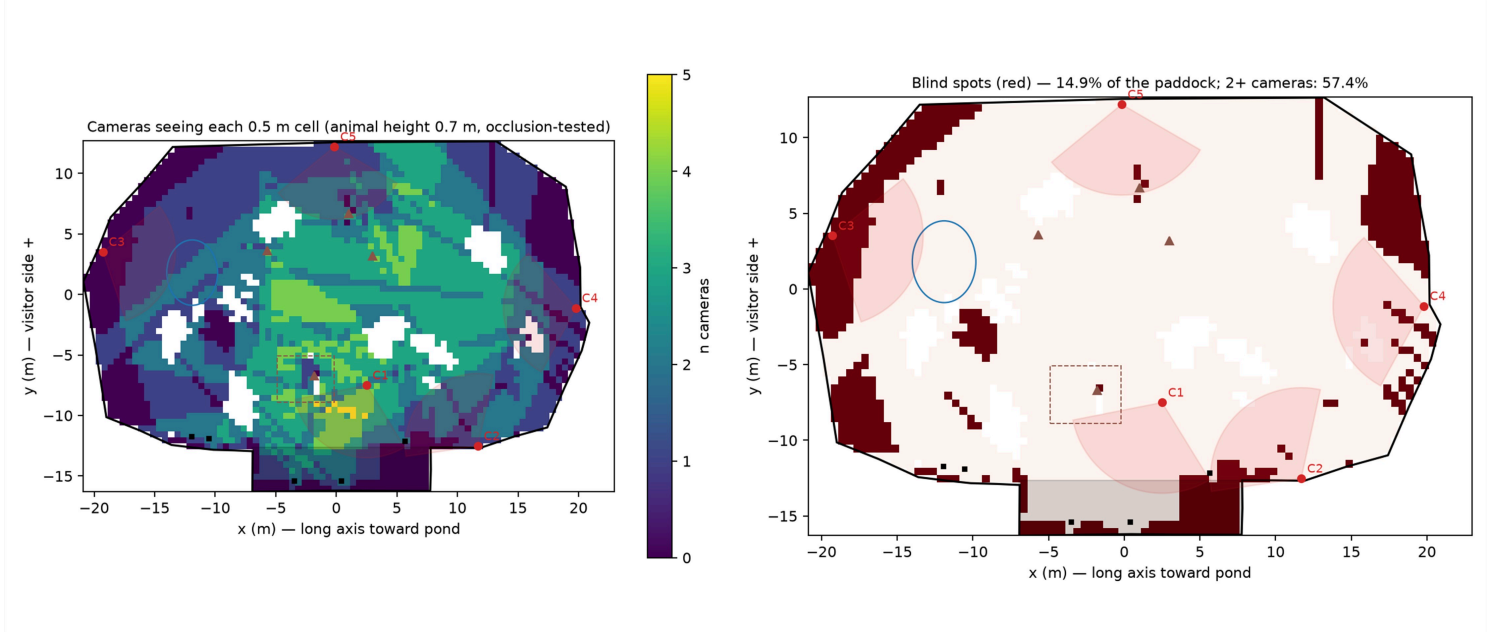


Figure 3

Design-stage coverage of the installed five-camera arrangement on a 0.5 m grid at bear height (0.7 m), occlusion-tested against the twin geometry with ground vegetation transparent. Left: number of cameras with unoccluded, detectable view of each cell. Right: blind cells (red), 14.9% of the paddock. Camera positions and fields of view in red; pond ellipse in blue; den pocket shaded.

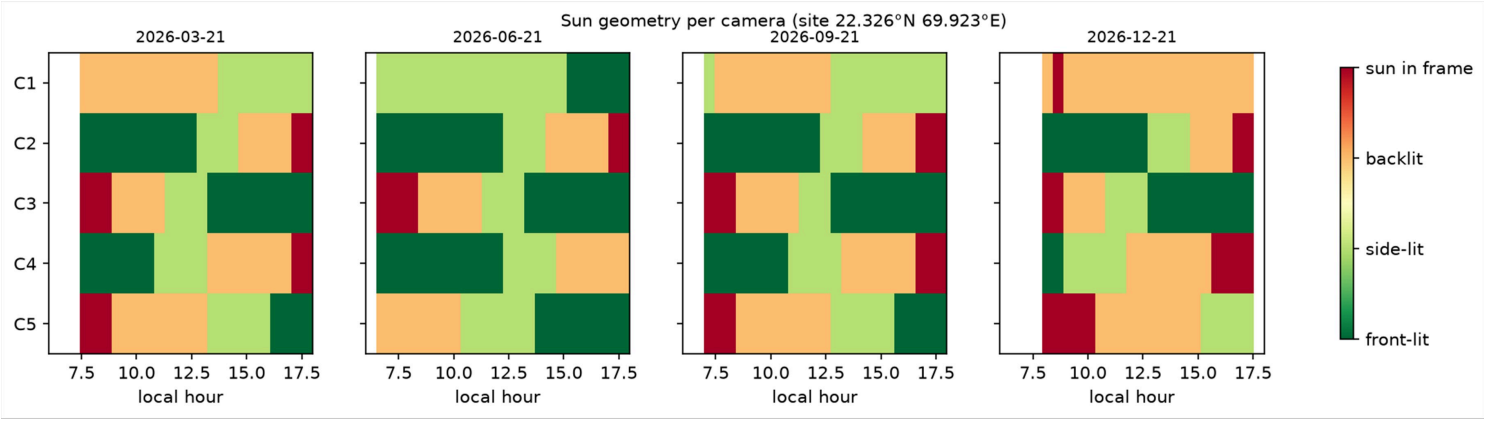


Figure 4

Predicted solar exposure per camera and local hour for nominal equinox and solstice dates, computed from the site's solar geometry (22.326°N, 69.923°E): front-lit, side-lit, backlit, and sun-in-frame conditions. The model predicts up to 10 h d⁻¹ of degraded (backlit or sun-in-frame) conditions for the south-facing den-door camera C1 at the winter solstice, and 7.5 h for the visitors-gallery camera C5.

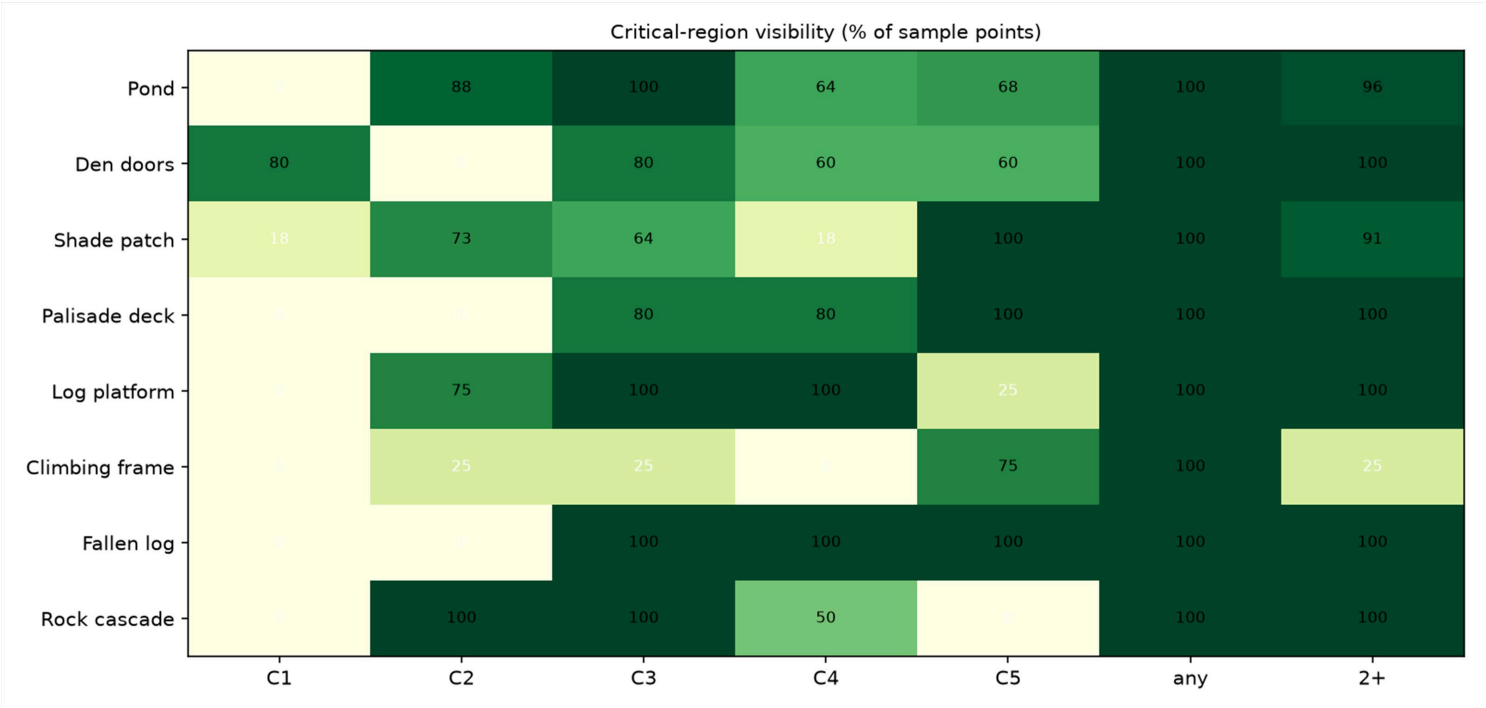


Figure 5

Critical-region visibility matrix (% of sampled points per region visible per camera, by any camera, and by ≥2 cameras) for husbandry-relevant targets: pond, den doors, shade patch, enrichment structures, and rock cascade

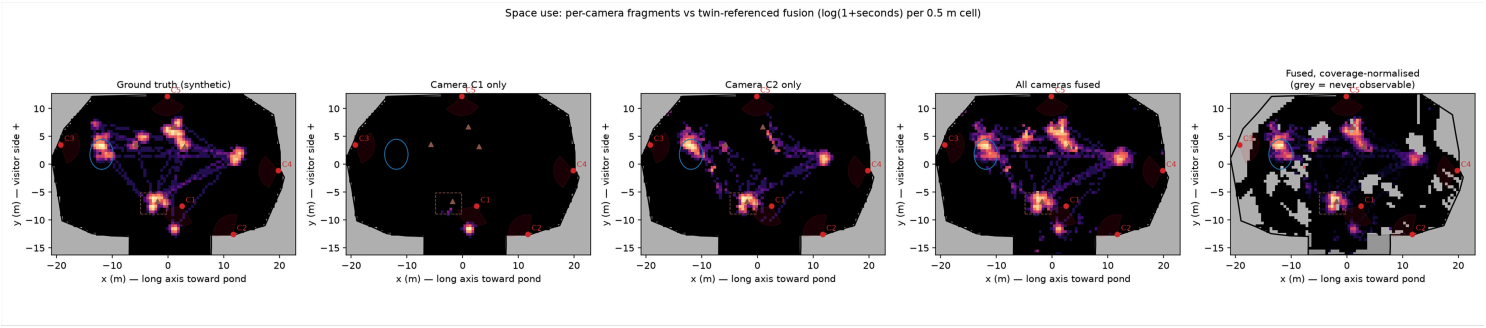


Figure 6

Space use for the one-hour validation scenario, log(1+seconds) per 0.5 m cell: ground truth; two single-camera fragments; the fused world-frame map; and the coverage-normalised map, in which never-observable cells are shown as unobservable (grey) rather than unused.

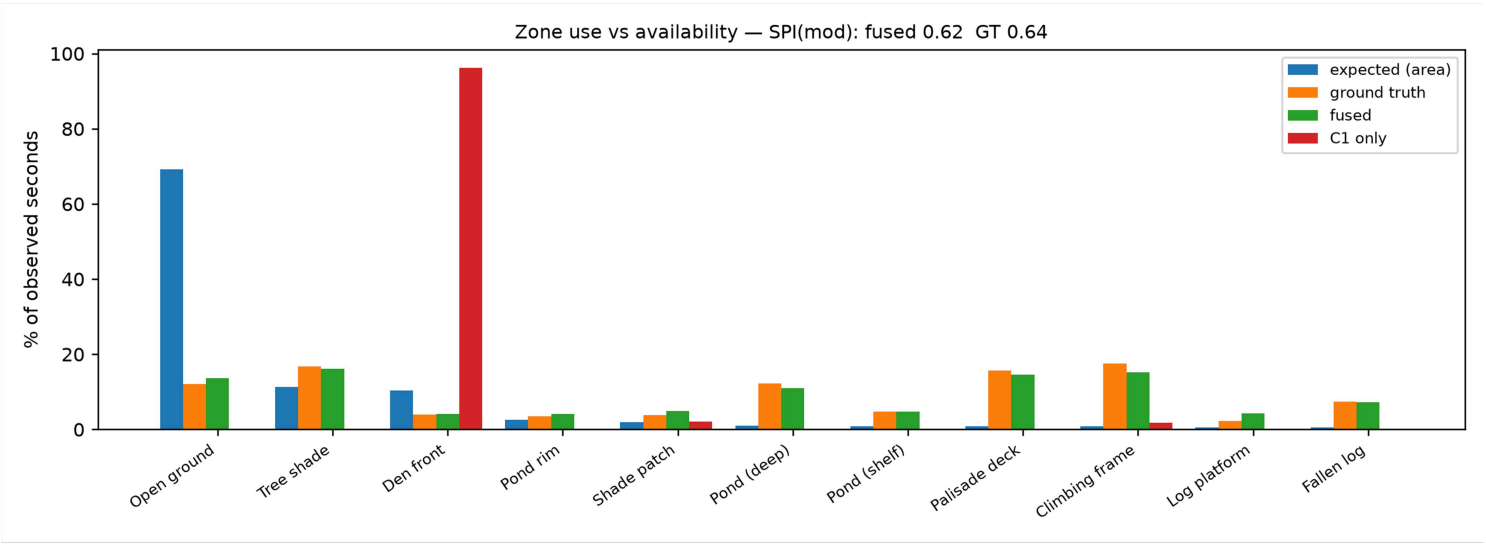


Figure 7

Zone use versus availability for the validation scenario: expected proportions from zone areas, ground truth, fused estimate, and a single-camera estimate; modified spread-of-participation index (SPI) 0.623 (fused) versus 0.640 (ground truth).

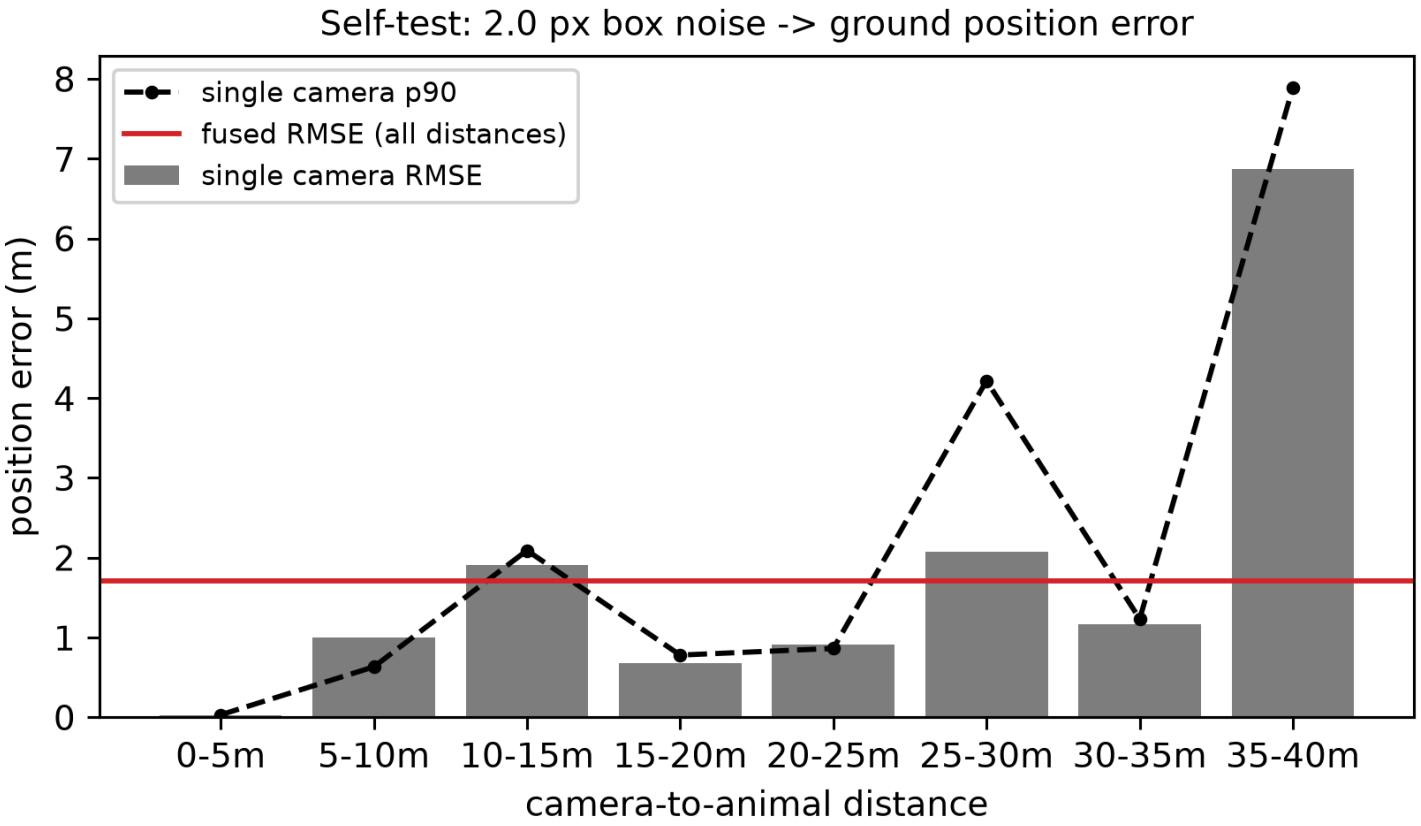


Figure 8

End-to-end position recovery in the self-test (2 px box noise): single-camera RMSE and 90th percentile by camera–animal distance, against the fused estimate. Multi-camera cluster estimates (61.7% of the 8,720 fused estimates, containing 81.2% of the underlying detections) recover 0.07 m median / 0.20 m RMSE.