RoboCap-10k: A Metrically Accurate Egocentric Dataset for Robot Learning

Kheri Hughes[1] Qasim Ali[1] Nikash Bhardwaj[1] Owen Akhibi[1]
Niresh Dravin[2] Vivek Sharma[2] Sam Cho[2] Ethan Lee[2] Paul Luo[2] Sunny Yu[2]
Jonathan Victor[2] Amanda Young[2] Michael Cho[2] Ademi Adeniji[1] Vincent Liu[1]

[1]Grounded Superintelligence [2]BitRobot

gsi.company

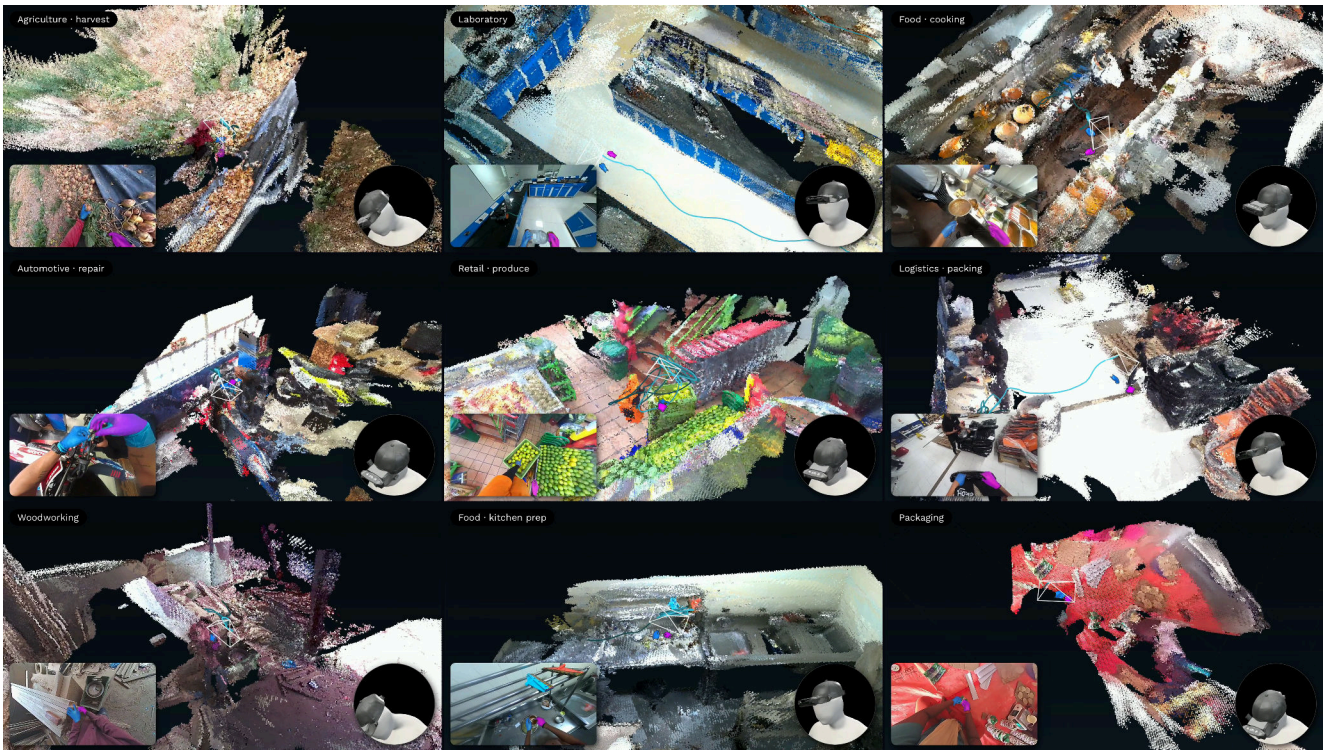


Figure 1. **RoboCap-10k**: 10,000 hours of egocentric occupational work, captured on calibrated RoboCap devices with metric 3D head and hand poses, and task and sub-task language annotated. Insets: egocentric view (bottom left) and head pose (bottom right).

Abstract

Egocentric video is the most scalable source of manipulation data for robot learning, but its difficulty lies in producing high-quality supervision. Existing large corpora lack metric 3D labels or rely on heterogeneous devices, motion-capture labs, or a headset's built-in real-time tracker, whose accuracy is not reported for the released data. We introduce RoboCap-10k: 10,000 curated hours of unscripted manipulation recorded during workers' paid shifts in 11 countries, spanning 21 industry verticals, 365 site types, and 619 tasks. Every hour is captured on one device, the RoboCap, a 250 g hat with six synchronized global-shutter cameras and dual IMUs. Online calibration brings every device below 1 px of epipolar error, visual-inertial SLAM yields gravity-aligned 6-DoF head trajectories, and a multi-view hand tracker yields world-frame cm-precise hand poses. Each episode is annotated with task and sub-task language captions.

1. Introduction

The pursuit of general-purpose robotics has exposed a data bottleneck: whereas foundation models for language and vision scale on internet text and imagery, robot policies remain starved for physical interaction data. Teleoperation provides demonstrations in the robot's own embodiment, but it is slow, expensive, and hard to extend to the long tail of environments and tasks. The field has therefore turned to egocentric human video, which captures fluid, purposeful manipulation across thousands of real environments at a fraction of the cost.

Egocentric video becomes action supervision only through estimation. Recent scaling studies show that pretraining on large egocentric corpora improves generalization of robot policies, and that 3D hand pose labels are needed to extract most of that benefit [4, 40, 55]. The quality of those labels is set by the capture hardware. The largest corpora are recorded with monocular head-mounted cameras [6, 23], from which metric scale can only be inferred from learned priors. Calibrated stereo rigs measure scale directly, but are sensitive to calibration: on the RoboCap, one pixel of disparity error corresponds to roughly one centimeter of depth at 0.65 m, comparable to the grasp tolerance of a parallel-jaw gripper [3]. Research headsets paired with optical motion capture reach millimeter accuracy, but only within instrumented volumes and with small participant pools [5, 27]. Federated efforts scale by pooling heterogeneous devices [46], which introduces cross-rig variation in both appearance and 3D estimation quality. Datasets labeled by a headset's own tracker inherit whatever that tracker outputs in real time, with no raw multi-view data to re-estimate from. RoboCap-10k instead releases the raw synchronized streams alongside offline labels whose accuracy is benchmarked (Section 5), so labels can be audited and regenerated as methods improve.

RoboCap-10k takes a different approach: it contains a single device model collecting real paid work processed by a single 3D stack. All 10,000 hours are recorded on the RoboCap [3], a 250 g hat whose camera layout is designed based on the requirements of near-field metric estimation, and all are processed by the same Grounded API pipeline with per-session online calibration. Rather than scripted household chores or tabletop tasks, the corpus records occupational work, including bread and pastry production, industrial garment stitching, engine and motorcycle repair, circuit-board soldering, rebar tying, and onion harvesting, in workplaces across India, Nigeria, Colombia, the United States, the United Kingdom, Mexico, Chile, and Spain.

Our contributions are:

1. **A 10,000-hour curated, single-device corpus of occupational manipulation.** Every hour is bimanual manipulation with a clear task and passes SLAM and hand-tracking quality checks. RoboCap-10k covers 21 industry verticals, 365 site types, and 619 distinct tasks in 11 countries across five continents. Because every hour comes from the same hardware design with per-unit factory calibration, there is no cross-device domain shift in either imagery or 3D label quality.
2. **Metric 3D annotations for every frame.** Online-calibrated, gravity-aligned 6-DoF head trajectories from multi-camera visual-inertial SLAM and world-frame MANO hand poses from a multi-view hand tracker, with per-session quality statistics.
3. **A two-tier language layer.** Interval-level sparse captions paired with dense, temporally segmented sub-task captions, each carrying scene, high-level task, and low-level action fields.

2. Related Work

2.1. Large-Scale Egocentric Datasets

Early egocentric corpora [12, 23] demonstrated the value of recording unscripted activity across thousands of hours and international cohorts, largely with head-mounted consumer cameras whose monocular, rolling-shutter footage does not support metric 3D reconstruction. Ego-Exo4D [24] and Nymeria [41] moved to calibrated Aria glasses with external cameras or body motion capture, HOI4D [38] used RGB-D for category-level hand–object interaction, and EgoDex [28] collected 829 hours of tabletop manipulation with on-device hand tracking from a mixed-reality headset.

A recent wave of datasets targets robot learning directly. Egocentric-10K [6] and its successors scale monocular factory footage to tens of thousands of hours or more, without 3D labels. EgoSuite-Open100K [36] has released the first 10,000 of a planned 100,000 hours of head- and wrist-camera footage across seven environment categories, including logistics and industry, with hand and body pose annotations and a license that permits commercial training. HumanNet [16] aggregates 1 million hours of human-centric video and recovers poses with monocular estimators. EgoVerse [46] federates 1,362 hours from heterogeneous devices through a shared processing pipeline. EgoLive [35] and Xperience-10M [53] use custom calibrated stereo rigs with hand and head annotations. RoboCap-10k differs from these in three respects (Table 1): all data comes from a single device model with two calibrated stereo baselines plus lateral coverage designed for the near field; calibration is re-estimated per session rather than assumed from the factory; and the content is occupational work across 21 industry verticals rather than daily-life or household routines.

Calibrated research devices [18, 39, 43] and multi-view hand–object benchmarks [5, 50] provide accurate ground truth, but depend on expensive motion-capture rigs, markerless capture cages, or the built-in real-time trackers of con-

Dataset	Released ego hours	Capture device	Single device	Calibrated stereo	Metric 6-DoF head	3D hands	Language	Occupational work	Setting
EPIC-KITCHENS-100 [12]	100	GoPro	✓	✗	✗	✗	✓	✗	Home kitchens
Ego4D [23]	3,670	Mixed consumer cameras	✗	✗	✗	✗	✓	✗	Daily life
Ego-Exo4D [24]	221	Aria + exo GoPros	✓	✗	✓	✗	✓	✗	Skilled activities
Nymeria [41]	300	Aria + XSens	✓	✗	✓	✗	✓	✗	Daily life
HOT3D [5]	14	Aria, Quest 3	✗	✗	✓	✓	✗	✗	Lab (mocap)
EgoDex [28]	829	Apple Vision Pro	✓	✗	✓	✓	✓	✗	Tabletop
EgoVerse [46]	1,362	Aria, iPhone, stereo rigs	✗	✗	✓	✓	✓	✗	Mixed
EgoLive [35]	1,680	Custom stereo head rig	✓	✓	✓	✓	✓	✗	Home, retail, service
Xperience-10M [53]	10,000	6-stream head rig + mocap	✓	✓	✓	✓	✓	✗	Daily life
EgoSuite-Open100K [36]	10,000	Head + wrist cameras	✗	✗	✗	✓	✗	✗	Home, retail, industry
Egocentric-10K [6]	10,000	Monocular head camera	✓	✗	✗	✗	✗	✓	Factories
RoboCap-10k (ours)	**10,000**	**RoboCap (6 cam, 2 IMU)**	✓	✓	✓	✓	✓	✓	**21 occupational verticals**

Table 1. **Comparison with egocentric datasets used for manipulation learning.** A property is marked only if it holds for, and is reported across, the full released dataset. "Occupational work" is counted if majority of the dataset is reported to be collected from individuals doing their occupational work. Hours count publicly released egocentric footage as of November 2026. RoboCap-10k is the only dataset combining calibrated stereo, metric head pose, and 3D hands with 10k+ released hours of occupational work.

sumer headsets [28], whose estimates cannot be re-run, inspected, or improved after capture, which confines them to tens of hours and small participant pools. Earlier wearable systems used downward-facing fisheye cameras on caps or body rigs for markerless body capture [1, 2, 49, 61], but targeted the problem of body pose estimation rather than near-field hand manipulation.

Portable demonstration interfaces and hand-mounted rigs [11, 56, 59] record in-the-wild interactions with instrumented end-effectors or gloves, at the cost of altering natural hand use. The RoboCap leaves the hands bare.

2.2. Egocentric Visual-Inertial SLAM

Head-worn ego-motion estimation must cope with rapid rotation, close-range parallax, and the wearer's own hands occluding the view [14, 32]. Classical visual-inertial systems [8, 22, 33, 34, 47, 54, 58] combine on-manifold IMU preintegration [19, 20] with sliding-window optimization [15]. GPU implementations [31] extend these to multi-camera rigs, learned dense pipelines [42, 57, 60] trade sparse optimization for dense correspondence, and Gaussian-splatting backends [9, 68] improve offline mapping fidelity. All remain vulnerable when manipulation dominates the forward field of view.

Wearable rigs also deform. Offline toolkits [21] recover a rigid calibration at the factory, but mechanical and thermal stress in the field shift it. Across the RoboCap fleet, the relative rotation between stereo cameras drifts by 0.2–1.1° (median ≈0.5°), producing 1–5 px of epipolar residual and a 1–2 cm depth bias at manipulation range [3]. Continuous stereo self-calibration has been studied for vehicles [13]; we apply per-session rotational self-calibration to every recording in the corpus (Section 3.2).

2.3. Egocentric Hand Pose Estimation

Parametric hand models [51] and transformer-based monocular reconstructors [44, 45, 52, 66] have made single-image 3D hand recovery robust in the wild, and world-space extensions recover trajectories from moving cameras [65, 67]. Monocular methods, however, regress depth under a weak-perspective assumption from apparent hand size, so their outputs are metrically ambiguous and inconsistent across views. Interaction benchmarks [10, 25, 62] provide accurate multi-view supervision but are captured in controlled settings. Multi-view and learnable-triangulation methods [26, 27, 29, 63, 64] resolve the depth ambiguity by fusing synchronized views; production systems of this kind derive labels from motion capture. The Grounded API hand tracker is a feed-forward multi-view model trained on in-the-wild RoboCap recordings labeled mainly by offline multi-view optimization rather than motion capture [3].

2.4. Egocentric Robot Learning

Imitation learning from first-person human video [30, 37] and cross-embodiment pretraining on human motion [4, 40, 48] show that human demonstrations transfer to robots once hand and head poses are mapped to a shared action space. Because retargeting consumes estimated poses, estimation error enters the policy's training distribution as label noise. RoboCap-10k is designed to reduce that noise at scale while keeping the raw multi-view streams, so labels can be regenerated as estimators improve.

3. Preliminaries

We summarize the RoboCap hardware and Grounded API processing pipeline; full details and evaluations are in the companion report [3].

Subsystem	Optics / FoV	Baseline / Overlap	Rate / Sync
Front stereo	Fisheye $180^\circ \times 112^\circ$	86.8 mm (157° overlap)	1080p @ 30 Hz HW trigger
Eye stereo	Fisheye $180^\circ \times 112^\circ$	106.2 mm (154° overlap)	1080p @ 30 Hz HW trigger
Lateral ($2\times$)	Wide-angle $161^\circ \times 98^\circ$	85° off-axis, 30° down	1080p @ 30 Hz HW trigger
IMUs ($2\times$)	6-axis, crown sides	Cam–IMU $\leq$3 ms	200 Hz
Clock	24 MHz TCXO	1-to-6 PWM	$\pm$1.81 ppm

Table 2. RoboCap sensor suite (adapted from [3]). All cameras are global-shutter, 10-bit RGB with $f \approx 590$–630 px.

3.1. The RoboCap

The RoboCap integrates its sensor payload into a 250 g hat that distributes weight over the crown, avoiding the nose-bridge load of glasses during multi-hour sessions (Table 2). Six global-shutter 1920×1080 cameras run at 30 Hz: a front and an eye stereo pair with fisheye optics give 157° and 154° of binocular overlap on 86.8 mm and 106.2 mm baselines, keeping the hands inside a metric stereo frustum across the reach envelope, and two lateral wide-angle cameras, mounted 85° off the forward axis and pitched 30° down, capture reaches that leave the stereo field and keep the trajectory solver constrained when the hands cover the front pair. All cameras are exposure-triggered by a single PWM signal fanned out 1-to-6, so frames are simultaneous by construction. Two 6-axis IMUs on opposite sides of the crown sample at 200 Hz in the same 24 MHz TCXO clock domain, with camera–IMU offsets identified per unit to a residual of at most 3 ms. Streams are encoded on-device (H.265, 4 Mbps per camera) to 214.5 GB of onboard storage. A tethered 10,000 mAh bank sustains roughly six hours of recording.

3.2. Factory and Online Calibration

Every unit is calibrated at the factory against an AprilGrid target with continuous-time batch optimization [21], recovering pinhole intrinsics with 4-parameter Kannala–Brandt distortion, camera–camera and camera–IMU extrinsics, and time offsets. Units outside per-parameter tolerance gates are reworked before shipment.

Field drift is dominated by rotation [3]. We therefore re-estimate the rotations of all six cameras jointly for each recording session, holding every factory translation fixed: epipolar constraints cannot observe baseline length, so metric scale stays anchored to the factory calibration. This extends the pairwise procedure of [3], which we revised after observing in-field deformation of deployed units. The left-front camera is the reference and is never moved; each other camera c receives a rotation $\mathbf{R}_c \in SO(3)$ relative to it. For every camera pair $(a, b) \in \mathcal{P}$ with overlapping fields of view (the front and eye stereo pairs, the front–eye pairs, and each lateral camera with the stereo cameras it overlaps), correspondences $\mathcal{M}_{ab}$ are lifted to normalized image coordinates $\hat{\mathbf{x}} = \pi^{-1}(\mathbf{u})$ through the Kannala–Brandt model, and we solve

$$\min_{\{\mathbf{R}_c\}} \sum_{(a,b)\in\mathcal{P}} \sum_{i\in\mathcal{M}_{ab}} \rho_\delta\left(\frac{(\hat{\mathbf{x}}_i'^{\top}\mathbf{E}_{ab}\hat{\mathbf{x}}_i)^2}{\|(\mathbf{E}_{ab}\hat{\mathbf{x}}_i)_{1:2}\|^2 + \|(\mathbf{E}_{ab}^{\top}\hat{\mathbf{x}}_i')_{1:2}\|^2}\right), \tag{1}$$

where $\mathbf{E}_{ab} = [\mathbf{t}^0_{ab}]_\times \mathbf{R}_b \mathbf{R}^0_{ab} \mathbf{R}_a^{\top}$ composes each camera's correction with the factory relative pose $(\mathbf{R}^0_{ab}, \mathbf{t}^0_{ab})$, and ρ_δ is a Huber kernel. Solving jointly keeps the corrections consistent across the rig: a camera that participates in several pairs receives a single rotation, so pairwise estimates cannot disagree. A session is rejected if any of the six cameras is degenerate, or if the feature-matching reprojection error after correction exceeds 1 px on the front cameras or 1.5 px on the side cameras.

Across the corpus, online calibration brings every device below 1 px of average Sampson error, with a median device of 0.63 px on the front stereo pair and 0.60 px on the eye pair (Section 5.2); devices that cannot be calibrated below 1 px are excluded from the release.

3.3. Visual-Inertial SLAM

Head trajectories are estimated in a metric, gravity-aligned world frame by Grounded API's multi-camera visual-inertial SLAM [3]. A GPU-accelerated patch-based optical-flow front end feeds a square-root sliding-window estimator [15, 58] with preintegrated inertial factors [19]. On the RoboCap the front pair forms a stereo rig that gives landmarks metric depth from their first observation, while the two lateral cameras (four cameras in total; the eye pair is not used for SLAM) track their own landmarks monocularly, keeping the estimator constrained when hands and tools cover the front views. Appearance-based place recognition detects revisits, and the released trajectories come from an offline keyframe bundle adjustment over the whole session once all revisits are known, which is roughly twice as accurate as the causal estimate available during capture. One configuration is used for every session.

3.4. Hand Pose Estimation

RoboCap-10k provides per-frame world-frame hand poses for both hands as MANO [51] parameters together with 21 keypoints (16 joints plus 5 fingertips). The Grounded API tracker is a feed-forward multi-view ViT that takes hand crops from all cameras that see the hand, together with their calibrated intrinsics and extrinsics, and predicts a single MANO pose consistent across views by construction. It is trained on a large in-the-wild RoboCap corpus whose labels come from an offline multi-view optimization that bootstraps monocular priors [44, 45] and refines a single MANO fit per trajectory for reprojected appearance and

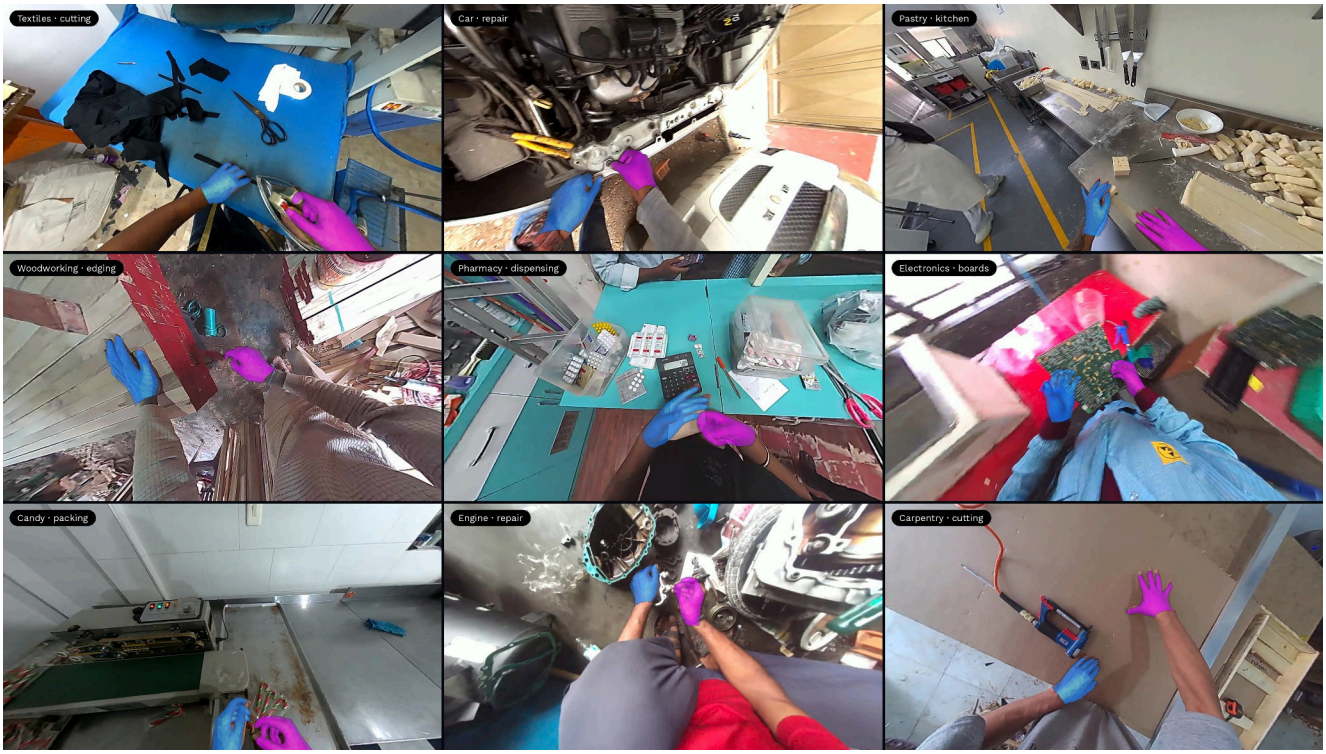


Figure 2. **Hand pose estimates across workplaces.** Egocentric frames from the nine sessions of Figure 1, with the estimated MANO hand meshes rendered over the image (blue: left hand, magenta: right hand). The tracker holds up through gloves, tool and object contact, partial occlusion, clutter and outdoor lighting. Appendix B shows 36 further sessions.

plausibility. The training data also includes a small motion-capture split recorded on the RoboCap (Section 5.4). This gives robustness to motion blur, tool occlusion, gloves, and difficult lighting. Figure 2 shows the resulting hand meshes across workplaces. Section 5.4 summarizes its accuracy.

4. The RoboCap-10k Corpus

RoboCap-10k comprises 10,000 hours (1,363,915 episodes) of unscripted egocentric manipulation recorded during real work. An episode is one continuous clip of at least 10 s in which both hands are tracked and SLAM is available; episodes run from 10 s to 600 s (median 17.8 s, mean 26.4 s, 90% under 48.7 s). These are curated hours, not raw recording time: each is retained only if it shows active bimanual manipulation with an identifiable task and passes the calibration, SLAM, and hand-tracking checks in Section 5. The corpus totals about 1.08×10^9 synchronized timesteps and 6.5×10^9 camera frames across the six streams. Hours are wall-clock capture time, and each hour contains six synchronized video streams. Figure 3a shows how the hours are distributed across countries.

4.1. Design Principles

We design RoboCap-10k around four principles.

Environment diversity. Policies degrade sharply under visual distribution shift. RoboCap-10k records activity in 365 types of real workplaces, from bakeries and garment factories to car repair garages, PCB assembly lines, and clinical diagnostic labs, with their natural clutter, lighting, and tools left as they are.

Task diversity and hierarchy. Manipulation at work is bimanual, contact-rich, and long-horizon. The corpus covers 619 distinct tasks grouped into 14 manual skill groups, including tool use, assembly, deformable-object handling, and fine electronics work. Each episode pairs an interval-level description of intent with dense sub-task segments, supervising both task planning and short-horizon motor control.

Metric action labels. Learning control from uncalibrated video leaves models to solve an ill-posed inverse problem. RoboCap-10k pairs every synchronized frame with a gravity-aligned 6-DoF head pose and metric world-frame MANO hand poses, which together define the 3D end-effector trajectories used to retarget human motion to robots.

Industry vertical	Site types	Tasks	Easy	Medium	Hard	Skill groups	Countries
Food & Beverage	39	117	43	70	4	4	4
Furniture & Woodworking	26	68	1	30	37	4	3
Automotive & Transport	36	66	3	30	33	2	6
Industrial Manufacturing	39	62	25	25	12	4	4
Agriculture & Farming	32	56	16	39	1	1	3
Clothing & Fashion	34	49	5	19	25	2	4
Construction & Hardware	33	43	1	22	20	1	4
Household / Domestic	15	25	19	6	0	4	4
Logistics & Warehousing	9	22	21	1	0	1	3
Retail & Consumer Goods	20	20	19	1	0	2	4
Creative Workshops & Crafts	13	20	3	13	4	1	4
Printing, Packaging & Design	14	18	6	7	5	3	2
Beauty & Personal Care	13	12	0	2	10	2	3
Cleaning, Laundry & Sanitation	13	11	4	7	0	2	2
Repair Services	9	11	0	0	11	3	2
Healthcare & Pharmacy	6	4	0	2	2	2	2
Laboratory / Scientific	4	4	0	0	4	2	1
Energy & Utilities	3	3	0	0	3	3	2
Administrative & Office	3	2	2	0	0	1	1
Hospitality & Lodging	3	0	0	0	0	0	1
Sports & Recreation	2	0	0	0	0	0	1
Other / Mixed Trades	5	6	2	3	1	3	4
Total (distinct)	**365**	**619**	**170**	**277**	**172**	**14**	**11**

Table 3. **RoboCap-10k diversity by industry vertical**, sorted by number of distinct tasks; shading of the count columns is proportional to each column's maximum. Other / Mixed Trades groups work that does not sit in one industry and is not counted as a vertical. Hospitality and Sports have footage on file awaiting task labeling. Totals count distinct labels: site-type names repeat across verticals, so the column sums to 369 but totals 365 (371 vertical–site pairs). Country counts include Spain (one Logistics site).

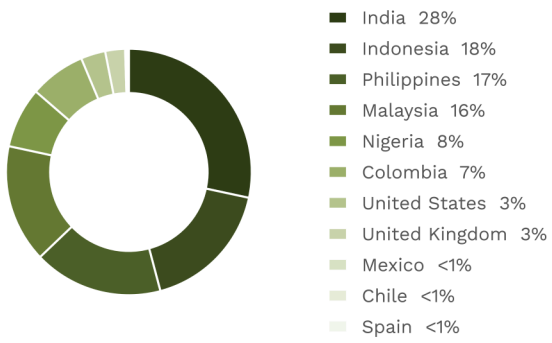
Standardized hardware and calibration. All data comes from one device model [3], so imagery and 3D label quality are consistent across the corpus. Per-session online calibration removes the extrinsic drift that accumulates in deployed units, without targets or operator intervention.

4.2. Collection Protocol

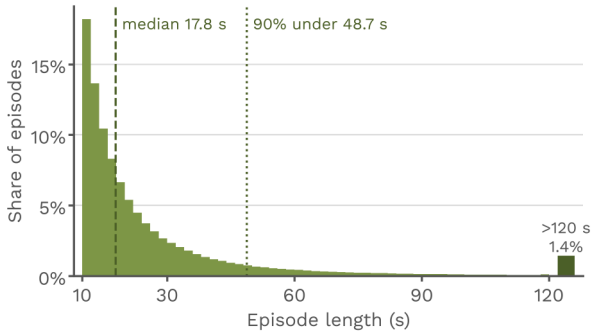
Data was collected through partner enterprises in 11 countries. Wearers are existing employees who record during their regular paid shifts of 6–12 h; the cap is the only change to their work, and tasks are not scripted, staged, or repeated for the camera. A local lead in each country recruits partner enterprises and coordinates the staff who rotate devices and train new wearers. Each unit shipped with its own factory calibration. Caps stay on site; every 1–3 days their recordings are offloaded to storage media, transported to a regional data center, and uploaded to the cloud, where Grounded API processes each session independently.

After processing, footage is curated into the released corpus at the level of episodes: contiguous segments in which one person performs one recognizable task in one setting. Trained annotators segment recordings into episodes following written guidelines that define task boundaries and exclude unusable footage (transit between work points, idle time, obstructed or unstable views), and every segmentation is reviewed by a second person; segmentations that fail review are returned for revision rather than silently corrected. A label is counted only if it was entered by an annotator, confirmed by a reviewer, or produced by a labeling stage validated against human annotation, and episodes missing any required label are held back. Labels are drawn from a centrally maintained, versioned taxonomy; automated checks flag implausible combinations, such as a site type that cannot occur in the assigned vertical, and synonymous labels are merged. Each episode is further checked for complete, consistently timed camera streams and is deduplicated against all other episodes by file fingerprint, content signature, and recording metadata.

Every session is labeled with an industry vertical, a site type (the kind of workplace), and a task (the activity in view). Labels come from two sources: the recording metadata, and a frame-level review of the footage performed

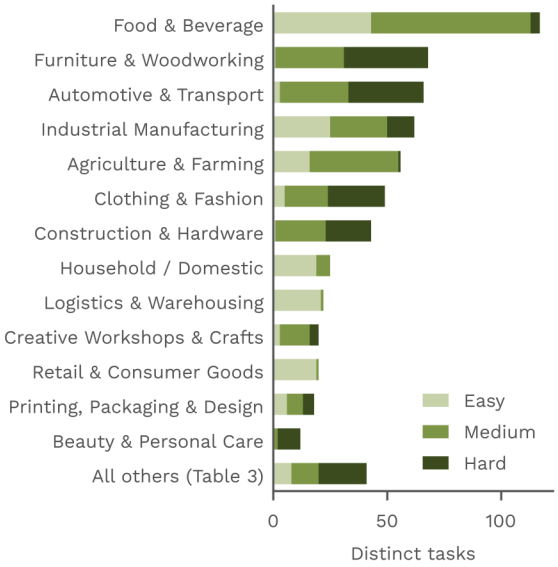


(a) Released hours by country.



(b) Episode length distribution.

Figure 3. **Corpus statistics.** (a) Share of released hours by country. (b) Share of episodes by length in 2 s bins; episodes longer than 120 s are grouped in the final bar.



(a) Distinct tasks per industry vertical.



(b) Share of distinct tasks per skill group.

Figure 4. **RoboCap-10k taxonomy.** Every task is labeled with the industry *vertical* it is recorded in (a) and, independently, the *skill group* of manual work it requires (b); both are broken down by difficulty. (a) shows the 13 verticals with the most tasks; the remaining six and Other / Mixed Trades are combined in the last row and listed individually in Table 3.

blind to that metadata. Where both exist, the footage review takes precedence. Labeling is continuous, so counts grow as review proceeds; the figures in this paper are a snapshot. No audio is recorded.

Consent and ethics. Every recorded operator signed an informed consent form, and every partner enterprise signed a data agreement before recording began; participation was voluntary. Wearers are compensated on top of their regular wages and footage is never used to evaluate individual workers' performance. Compliance with local labor and privacy law is the responsibility of the partner operating in each country. No audio is captured, faces are anonymized in all released video (Section 4.6), and released metadata identifies only the country and site type, never the site, employer, or participant.

4.3. Dataset Taxonomy

RoboCap-10k is organized by a three-level taxonomy (Table 3): 21 industry verticals, plus an Other / Mixed Trades bucket for work that spans industries, contain 365 workplace site types, at which 619 distinct tasks are performed. Each task is assigned a difficulty tier and a manual skill group. The appendix lists every site type with its exam-

Category	Fields	Description
Links	`unit_id`, `parent_id`, `episode_id`	Joins each dense unit to its parent interval and sparse caption.
Timing	start/end frames and clip-relative seconds, `continuation_of_previous`	Clip-relative boundaries; flag for actions continuing from the preceding unit.
Text	`scene`, `high_level_task`, `low_level_task`, `full_caption`	Scene context, functional sub-task, and local action, joined into a three-sentence caption.
Grounding evidence	`evidence_frames`, `evidence_views`, `grounding_note`, `review_warnings`	Frame and view anchors supporting the caption, explanatory notes, and ambiguity flags.
Provenance	`origin`, `offline_edited`, `human_visual_review`	Generation path, post-hoc edits, and human review status.

Table 4. **Fields stored with each dense unit.** Every dense unit belongs to one interval, and every interval has one sparse caption; the **Links** fields encode that hierarchy. Only the **Text** fields form the caption itself; the rest is metadata.

One interval: its sparse caption and three dense units		**One dense unit**: all fields of unit 3 (bold, left)	
Sparse, 0.00–14.80 s	*Tears a piece of paper towel from a roll, folds it, and wipes a glass beaker.*	Start / end	9.50–14.80 s (frames 285–444), clip-relative
Dense, 0.00–3.50 s	*Tears a section of paper towel from the roll.*	`scene`	*A person is standing at a laboratory workbench handling paper towels and a small container.*
Dense, 3.50–9.50 s	*Folds the torn paper towel sheet.*	`high_level_task`	*Cleans a small container using a folded paper towel.*
Dense, 9.50–14.80 s	***Wipes the interior of a small clear container with the folded paper.***	`low_level_task`	*Wipes the interior of a small clear container with the folded paper.*

Table 5. **The two caption tiers on one interval.** *Left:* one sparse caption describes the whole interval, and dense units split it into consecutive segments; each segment is listed here by its `low_level_task`. *Right:* the fields stored for one dense unit (unit 3, bold on the left): its time span, a `scene` description, the sub-task it belongs to (`high_level_task`), and the physical action (`low_level_task`). The three text fields are joined into the unit's `full_caption`.

ple tasks (Table 13); the full task list is released with the dataset. Difficulty is rated per task:

1. *Easy* tasks are single-step and need no training (170 tasks, e.g., sorting groundnuts, picking products from shelves).
2. *Medium* tasks are multi-step and need coordination or planning (277, e.g., shaping dough into loaves, portioning meals into trays).
3. *Hard* tasks are skilled work requiring training, experience, or domain expertise, including manual crafts (172, e.g., soldering components onto PCB panels, tying rebar cages, weaving cane onto chair frames, stitching garments on industrial machines).

Skill groups describe the kind of manual skill performed independent of industry. Counts are of distinct labels after normalizing case and spacing; where one label covers several workflows the count is conservative, so these figures are a floor on true variety. The 14 skill groups cut across verticals (Figure 4b): Food Preparation & Cooking covers 109 tasks across Food & Beverage and Household, while Organization & Inventory appears in nine verticals (Table 11).

The distribution is long-tailed by design (Figure 4a). The four largest verticals (Food & Beverage, Furniture & Woodworking, Automotive & Transport, and Industrial Manufacturing) together account for 313 of the 619 tasks, while specialized verticals such as Laboratory / Scientific and Energy & Utilities contribute a handful of hard, high-precision tasks each.

4.4. Sparse and Dense Language Captions

To support language-conditioned policies and world models, RoboCap-10k provides two tiers of language. Rather than slot-filling a fixed taxonomy of objects and verbs, the captions describe what happens in free text and link fine-grained action segments to interval-level intent. We use three terms throughout. An *interval* is a contiguous demonstration clip within an episode. Each interval has exactly one *sparse caption*: a single free-text sentence describing the whole clip. Each interval is also cut into consecutive *dense units*: shorter time segments, each with its own structured caption made of the text fields below. Table 5 shows one interval with both tiers.

Tier 1: sparse captions (one per interval). Each demonstration interval carries one high-level caption of the overall task and object interaction (e.g., *"Tears a piece of paper towel from a roll, folds it, and wipes a glass beaker"* over 0.00–14.80 s). This caption expresses the global intent of the clip and serves as the anchor for dense annotations.

Tier 2: dense units (many per interval). Each interval is temporally segmented into non-overlapping dense units,

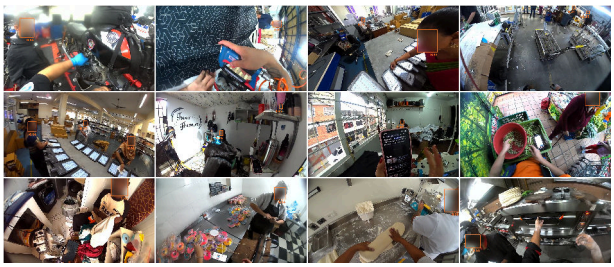


Figure 5. **Face anonymization example** (from a front camera). Our face blurring model identifies faces across all varied environments, demographics, camera models, and orientations in camera frame. See Figure 9 for more examples.

	Ours		RetinaFace	
False positives / frame	0.2	0.05	0.2	0.05
R / P, IoU 0.3	87.7 / 77.4	77.3 / 92.4	29.8 / 54.0	25.1 / 79.6
R / P, IoU 0.5	82.0 / 71.6	73.4 / 87.4	28.5 / 51.7	24.6 / 78.1
Recall on the hardest cases				
Small faces (24–32 px)	69.8	51.1	6.5	1.4
Eye cameras	88.8	78.9	2.5	1.7
Image periphery	85.1	71.2	18.1	15.8

Table 6. **Face detection on held-out sessions** (1,336 faces). Detectors are compared at thresholds giving 0.2 and 0.05 false positives per frame. Hardest-case rows report recall at IoU 0.3 for the smallest faces, the eye cameras, and the outer quarter of the image.

each covering one sub-task or action, with clip-relative start and end times in seconds and frames. A dense unit is not a separate caption type; it is a time segment, and its caption is spread across three text fields:

- `scene`: the workspace and setting (e.g., *"A person is standing at a laboratory workbench handling paper towels and a small container"*);
- `high_level_task`: the sub-task being performed (e.g., *"Cleans a small container using a folded paper towel"*);
- `low_level_task`: the local physical action (e.g., *"Wipes the interior of a small clear container with the folded paper"*).

The three fields are joined into a three-sentence `full_caption`, with no minimum length. Dense units are generated by Gemini 3.8 Flash from the two front-stereo views, which keep the near field in view. The sparse caption is retained for linking but is not given to the model, so the two tiers are produced independently. Each unit also stores the grounding and provenance fields in Table 4. In short: the sparse caption answers *what is this clip about*, and the dense units answer *what is happening at each moment*, at two levels of detail (`high_level_task` for the sub-task, `low_level_task` for the physical action).

4.5. Format and Release

RoboCap-10k is released under the Creative Commons Attribution 4.0 International license (CC BY 4.0). Every episode is provided in two formats. The LeRobot format [7] is intended for training: each camera is stored as an H.265-encoded MP4, and per-frame data (the SLAM head pose, MANO parameters and keypoints for both hands, and IMU readings) is stored in Parquet tables indexed by frame, so the corpus plugs directly into existing LeRobot dataloaders and policy-training code. A Rerun recording (`.rrd`) of the same episode is intended for inspection: it places the six camera streams, head trajectory, and hand meshes in a shared 3D world frame on a common timeline. Each episode also carries its factory and online-corrected intrinsics and extrinsics, and a JSON sidecar holding taxonomy labels, sparse and dense captions, and per-session quality statistics. All streams share one device clock, so video frames, poses, and captions correspond one-to-one by frame index. At the recorded bitrate the six video streams total roughly 108 TB.

4.6. Face Anonymization

Egocentric workplace footage routinely contains customers, co-workers, passersby, and occasionally the wearer in mirrors or reflective surfaces. Off-the-shelf face detectors are trained on eye-level photographs and miss many faces in this setting. For example, faces are seen from above through fisheye optics at the image periphery where distortion is strongest. Faces can also be small and distant, partially occluded by tools and hands, or blurred by head motion. We therefore train our own detector on roughly 1,000 hours of RoboCap footage with human-labeled face boxes, sampled across all 11 countries, every camera position, and both indoor and outdoor sites.

The detector runs on every frame of all six streams. Its operating point balances two failure modes: a missed face is a privacy failure, while a false positive blurs part of the scene, and in egocentric footage that part is often the hands or the object being manipulated, which is exactly the signal the dataset exists to provide. We therefore release at 0.05 false positives per frame (Table 6), where precision is 92.4%. Detections are linked into tracks, propagated forward and backward along each track, and dilated before blurring, so a face missed in one frame is still covered by its neighbors. the dilation is 15% of box size, kept small so that blur does not spread onto nearby hands. Blurring is drawn as a Gaussian blur, and irreversible: released videos are re-encoded from the blurred frames and the originals are not distributed.

5. Annotation Quality

Absolute accuracy of Grounded API is established in the companion report against external benchmarks [3]; we sum-

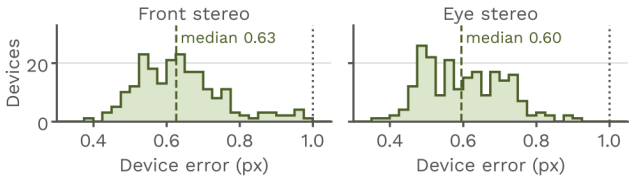


Figure 6. **Calibration error per device after online calibration.** Held-out Sampson error of each of the 216 devices in the release. Dashed lines mark the median; dotted lines mark 1 px.

	Front stereo (px)	Eye stereo (px)
Median	0.63	0.60
Mean $\pm\ \sigma$	0.64 ± 0.12	0.60 ± 0.11
P5–P95	0.48–0.88	0.46–0.77
Best device	0.40	0.37
Worst device	0.99	0.90

Table 7. **Distribution of per-device calibration error** (px) after online calibration, across the entire corpus. The spread is quasi-normal, slightly skewed towards lower errors.

Trajectories	Duration (h)	Median ATE (mm)	Mean ATE (mm)
4,669	153.4	3.0	3.19

Table 8. **Head trajectory accuracy** against motion capture on RoboMoCap [17]. We compute ATE after $\mathrm{SE}(3)$ alignment.

Benchmark	Metric	Error (mm)
RoboMoCap test [17]	PA-MPJPE	4.5
GSI-Internal [3]	MPJPE	9.2

Table 9. **Hand reconstruction accuracy** over 21 keypoints. RoboMoCap: 9.2 min of held-out motion capture across 4 sessions, fitted to MANO. GSI-Internal: simulated and motion-capture sequences with known hand shape, reported in [3].

marize it here and then report statistics computed over the corpus itself.

5.1. Natural Language Captions

Captions for 1,000 hours (10% of the corpus) were visually reviewed by human annotators against the source video. We fit our captioning pipeline to these labels and the resulting pipeline captions the entire dataset to standardize semantic ontology. Operators also provide basic metadata when they upload data to our services, including geography, vertical, and occupation, which is used to further ground our captioning pipeline.

5.2. Calibration Across the Corpus

We summarize calibration quality per device. For each of the 216 devices in the release, we average the held-out Sampson error after online calibration over all of its sessions, weighted by episode; the error of each camera includes all of its overlapping pairs, not only its stereo partner. Figure 6 and Table 7 show the resulting distribution across devices.

On the front stereo pair the median device error is 0.63 px (mean 0.64 px, standard deviation 0.12 px), and the eye pair is similar (median 0.60 px). Every device averages below 1 px on both pairs, with the worst at 0.99 px on the front pair and 0.90 px on the eye pair. Two devices whose front-pair error averaged above 1 px were excluded from the release, removing 34 hours (0.6%) of footage.

5.3. SLAM

The LaMAria benchmark [32] is a a head-worn visual-inertial benchmark of 63 city-scale test walks (27.5 h) scored against surveyed control points. GroundedSLAM, Grounded API's SLAM implementation, achieves an average control-point score of 72.4, nearly matching the performance of Aria's SLAM, which scores 75.9.

We further evaluate trajectory accuracy directly against optical motion capture on RoboMoCap [17], where a NOKOV system tracks a rigid body on the RoboCap headset at 120 Hz. We evaluate on 153.4 hours of motion capture data spread over the four headset units, recording dates, and seven activity families (locomotion, posture dynamics, loco-manipulation, whole-body environment interaction, dynamic recovery, static manipulation and long-horizon composites).

Each SLAM trajectory is metric and is evaluated as produced: the mocap clock offset is estimated from angular-speed correlation, the rigid-body-to-camera transform is solved per session with scale fixed to one, and the trajectory is aligned to the mocap with an $\mathrm{SE}(3)$ fit before computing the absolute trajectory error (ATE, position RMSE).

Table 8 summarises the results. On held-out sessions the median ATE is 3.0 mm and the mean is 3.19 mm. The metric scale is recovered to a median 0.17 % without any alignment scale, and the residual error is dominated by slow ($>$5 s) drift rather than frame-to-frame noise, whose high-frequency component is about 0.2 mm.

5.4. Hand Tracking

We evaluate hand poses against motion capture on RoboMoCap [17], concurrent work recorded on the RoboCap headset. RoboMoCap's training split (34.9 minutes) is part of the tracker's training data, its validation split is used for checkpoint selection, and we report results on its held-out test split (9.2 minutes). RoboMoCap does not provide the operator's hand shape, so absolute joint error (MPJPE)

would mix hand-size mismatch with tracking error. We therefore report Procrustes-aligned mean per-joint position error (PA-MPJPE) over the 21 hand keypoints, which removes global rotation, translation and scale and isolates finger articulation.

On the test split, Grounded API reaches a PA-MPJPE of 4.5 mm (Table 9), compared with 8.0 mm for a model trained without the RoboMoCap split. Because the test split contains only held-out sequences, this gain reflects learning RoboMoCap's joint and wrist conventions rather than fitting frame-level label noise, so most of the 8.0 mm is a convention mismatch rather than tracking error. For absolute error, we refer to the GSI-Internal benchmark of the companion report, a mix of simulated and motion-capture sequences where the hand shape is known, on which Grounded API reaches an MPJPE of 9.2 mm [3].

6. Limitations

Labels are estimates. All 3D annotations are produced by Grounded API, not measured. Calibration, SLAM and hand tracking are validated on external and RoboCap-specific benchmarks (Section 5), but millimeter ground truth cannot be collected in the wild, and hand errors concentrate where they matter most: fingers occluded at the contact interface during tight grasps. Language captions are model-generated; only about 10% of hours were reviewed by people.

Coverage and selection. Hours are spread over 11 countries, but India accounts for most verticals, and per-vertical labels are not yet complete for Indonesia, the Philippines and Malaysia. Curation keeps only bimanual manipulation that passes the 3D quality checks, so one-handed work, idle and transitional time, and the hardest tracking conditions (heavy occlusion, fast motion, poor lighting) are underrepresented relative to raw footage. Household and personal activities are largely absent, and task counts are uneven across verticals.

No touch or force. There are no tactile, shear or wrench measurements, so downstream policies must infer contact from vision alone.

Embodiment gap. Human hands have far more degrees of freedom than most robot end-effectors, and the human arms visible in the images differ from robot arms. Retargeting to grippers or dexterous hands must still handle joint limits, singularities and this visual gap.

7. Conclusion

RoboCap-10k pairs the scale of egocentric video with metric 3D labels by standardizing on one purpose-built device, re-estimating its calibration for every session, and processing every hour with the same SLAM and multi-view hand-tracking stack. The result is 10,000 hours of real occupational manipulation across 11 countries, 21 industry verticals and 619 distinct tasks, with world-frame head and hand trajectories and two tiers of language. We release it, with its raw multi-view streams, as pretraining data for generalist manipulation policies and world models, and as a testbed for a question the field has not yet answered: how much does the precision of human action labels matter at scale?

References

[1] Hiroyasu Akada, Jian Wang, Soshi Shimada, Masaki Takahashi, Christian Theobalt, and Vladislav Golyanik. UnrealEgo: A new dataset for robust egocentric 3D human motion capture. In *European Conference on Computer Vision (ECCV)*, pages 1–17, 2022. 3

[2] Hiroyasu Akada, Jian Wang, Vladislav Golyanik, and Christian Theobalt. 3d human pose perception from egocentric stereo videos, 2024. 3

[3] Qasim Ali, Nikash Bhardwaj, Kheri Hughes, Sam Cho, Ethan Lee, Paul Luo, Shinn Chong, Sunny Yu, Owen Akhibi, Alex Proshkin, Haotian Zhan, Niresh Dravin, Vivek Sharma, Jonathan Victor, Amanda Young, Michael Cho, Ademi Adeniji, and Vincent Liu. Robocap: A new platform for egocentric robot learning, 2026. 2, 3, 4, 6, 9, 10, 11

[4] Anonymous. Scaling dexterous manipulation with diverse egocentric human data. *arXiv preprint arXiv:2602.16710*, 2026. 2, 3

[5] Prithviraj Banerjee, Sindi Shkodrani, Pierre Moulon, Shreyas Hampali, Shangchen Han, Fan Zhang, Linguang Zhang, Jade Fountain, Edward Miller, Selen Basol, et al. Hot3d: Hand and object tracking in 3d from egocentric multi-view videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025. 2, 3

[6] Build AI. Egocentric-10k. https://huggingface.co/datasets/builddotai/Egocentric-10K, 2025. 2, 3

[7] Remi Cadene, Simon Alibert, Alexander Soare, Quentin Gallouedec, Adil Zouitine, Steven Palma, Pepijn Kooijmans, Michel Aractingi, Mustafa Shukor, Dana Aubakirova, Martino Russi, Francesco Capuano, Caroline Pascal, Jade Choghari, Jess Moss, and Thomas Wolf. LeRobot: State-of-the-art machine learning for real-world robotics in pytorch. https://github.com/huggingface/lerobot, 2024. 9

[8] Carlos Campos, Richard Elvira, Juan J. Gómez Rodríguez, José M. M. Montiel, and Juan D. Tardós. ORB-SLAM3: An accurate open-source library for visual, visual-inertial, and multimap SLAM. *IEEE Transactions on Robotics*, 37(6): 1874–1890, 2021. 3

[9] Zhihao Cao, Hanyu Wu, Li Wa Tang, Zizhou Luo, Wei Zhang, Marc Pollefeys, Zihan Zhu, and Martin R. Oswald. Mcgs-slam: A multi-camera slam framework using gaussian splatting for high-fidelity mapping, 2026. 3

[10] Yu-Wei Chao, Wei Yang, Yu Xiang, Pavlo Molchanov, Ankur Handa, Jonathan Tremblay, Yashraj S. Narang, Karl Van Wyk, Umar Iqbal, Stan Birchfield, Jan Kautz, and Dieter Fox. Dexycb: A benchmark for capturing hand grasping of objects, 2021. 3

[11] Cheng Chi, Zhenjia Xu, Chuer Pan, Eric Cousineau, Benjamin Burchfiel, Siyuan Feng, Russ Tedrake, and Shuran Song. Universal manipulation interface: In-the-wild robot teaching without in-the-wild robots. In *Robotics: Science and Systems (RSS)*, 2024. 3

[12] Dima Damen, Hazel Doughty, Giovanni Maria Farinella, Sanja Fidler, Antonino Furnari, Evangelos Kazakos, Davide Moltisanti, Jonathan Munro, Toby Perrett, Will Price, and Michael Wray. Scaling egocentric vision: The EPIC-KITCHENS dataset. In *European Conference on Computer Vision (ECCV)*, pages 720–736, 2018. 2, 3

[13] Thao Dang, Christian Hoffmann, and Christoph Stiller. Continuous stereo self-calibration by camera parameter tracking. *IEEE Transactions on Image Processing*, 18(7):1536–1550, 2009. 3

[14] Mateo de Mayo, Daniel Cremers, and Taihú Pire. The Monado SLAM dataset for egocentric visual-inertial tracking. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2025. 3

[15] Nikolaus Demmel, Christiane Sommer, Daniel Cremers, and Vladyslav Usenko. Square root marginalization for sliding-window bundle adjustment. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 13260–13268, 2021. 3, 4

[16] Yufan Deng and Daquan Zhou. HumanNet: Scaling human-centric video learning to one million hours. *arXiv preprint arXiv:2605.06747*, 2026. 2

[17] Mayur Deshmukh, Aaron Tung, Junyi Li, Nikash Bhardwaj, Ademi Adeniji, Vincent Liu, Michael Cho, and Jiafei Duan. Robomocap: A large-scale whole-body motion capture dataset with egocentric vision for robot learning. https://bitrobot-foundation.github.io/robomocap/, 2026. 10

[18] Jakob Engel, Kiran Somasundaram, Michael Goesele, Albert Sun, Alexander Gamino, Andrew Turner, Arjang Talattof, Arnie Yuan, Bilal Souti, Brighid Meredith, et al. Project aria: A new tool for egocentric multi-modal ai research. *arXiv preprint arXiv:2308.13561*, 2023. 2

[19] Christian Forster, Luca Carlone, Frank Dellaert, and Davide Scaramuzza. On-manifold preintegration for real-time visual-inertial odometry. *IEEE Transactions on Robotics*, 33 (1):1–21, 2017. 3, 4

[20] Christian Forster, Zichao Zhang, Michael Gassner, Manuel Werlberger, and Davide Scaramuzza. SVO: Semidirect visual odometry for monocular and multicamera systems. *IEEE Transactions on Robotics*, 33(2):249–265, 2017. 3

[21] Paul Furgale, Joern Rehder, and Roland Siegwart. Unified temporal and spatial calibration for multi-sensor systems. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1280–1286, 2013. 3, 4

[22] Patrick Geneva, Kevin Eckenhoff, Woosik Lee, Yulin Yang, and Guoquan Huang. OpenVINS: A research platform for visual-inertial estimation. In *IROS 2019 Workshop on Visual-Inertial Navigation: Challenges and Applications*, 2019. 3

[23] Kristen Grauman, Andrew Westbury, Eugene Byrne, Zachary Chavis, Antonino Furnari, Rohit Girdhar, Jackson Hamburger, Hao Jiang, Miao Liu, Xingyu Liu, et al. Ego4D: Around the world in 3,000 hours of egocentric video. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18995–19012, 2022. 2, 3

[24] Kristen Grauman, Andrew Westbury, Lorenzo Torresani, Kris Kitani, Jitendra Malik, Triantafyllos Afouras, Kumar Ashutosh, Vijay Baiyya, Siddhant Bansal, Bikram Boote, et al. Ego-Exo4D: Understanding skilled human activity from first- and third-person perspectives. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 19383–19400, 2024. 2, 3

[25] Shreyas Hampali, Sayan Deb Sarkar, and Vincent Lepetit. Ho-3dv3: Improving the accuracy of hand-object annotations of the ho-3d dataset, 2021. 3

[26] Shangchen Han, Beibei Liu, Randi Cabezas, Christopher D. Twigg, Peizhao Zhang, Jeff Petkau, Tsz-Ho Yu, Chun-Jung Tai, Muzaffer Akbay, Zheng Wang, et al. MEgATrack: Monochrome egocentric articulated hand-tracking for virtual reality. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 39(4):87:1–87:13, 2020. 3

[27] Shangchen Han, Po-Chen Wu, Yubo Zhang, Beibei Liu, Linguang Zhang, Zheng Wang, Weiguang Si, Peizhao Zhang, Yujun Cai, Tomas Hodan, et al. UmeTrack: Unified multi-view end-to-end hand tracking for VR. In *SIGGRAPH Asia 2022 Conference Papers*, pages 1–9, 2022. 2, 3

[28] Ryan Hoque, Peide Huang, David J. Yoon, Mouli Sivapurapu, and Jian Zhang. EgoDex: Learning dexterous manipulation from large-scale egocentric video. *arXiv preprint arXiv:2505.11709*, 2025. 2, 3

[29] Karim Iskakov, Egor Burkov, Victor Lempitsky, and Yury Malkov. Learnable triangulation of human pose. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 7718–7727, 2019. 3

[30] Simar Kareer, Dhruv Patel, Ryan Punamiya, Pranay Mathur, Shuo Cheng, Chen Wang, Judy Hoffman, and Danfei Xu. EgoMimic: Scaling imitation learning via egocentric video. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2025. arXiv:2410.24221. 3

[31] Alexander Korovko, Dmitry Slepichev, Alexander Efitorov, Aigul Dzhumamuratova, Viktor Kuznetsov, Hesam Rabeti, Joydeep Biswas, and Soha Pouya. cuVSLAM: CUDA accelerated visual odometry and mapping. *arXiv preprint arXiv:2506.04359*, 2025. 3

[32] Anusha Krishnan, Shaohui Liu, Paul-Edouard Sarlin, Oscar Gentilhomme, David Caruso, Maurizio Monge, Richard Newcombe, Jakob Engel, and Marc Pollefeys. Benchmarking egocentric visual-inertial slam at city scale, 2025. 3, 10

[33] Stefan Leutenegger. OKVIS2: Realtime scalable visual-inertial SLAM with loop closure. *arXiv preprint arXiv:2202.09199*, 2022. 3

[34] Stefan Leutenegger, Simon Lynen, Michael Bosse, Roland Siegwart, and Paul Furgale. Keyframe-based visual-inertial odometry using nonlinear optimization. *The International Journal of Robotics Research*, 34(3):314–334, 2015. 3

[35] Yihang Li, Xuelong Wei, Jingzhou Luo, et al. EgoLive: A large-scale egocentric dataset from real-world human tasks. *arXiv preprint arXiv:2604.23570*, 2026. 2, 3

[36] Lightwheel and Hugging Face. EgoSuite-Open100K: 100,000 hours of egocentric human data for physical AI. https://huggingface.co/blog/egosuite-open100k, 2026. 2, 3

[37] Vincent Liu, Ademi Adeniji, Haotian Zhan, Siddhant Haldar, Raunaq Bhirangi, Pieter Abbeel, and Lerrel Pinto. EgoZero: Robot learning from smart glasses. *arXiv preprint arXiv:2505.20290*, 2025. 3

[38] Yunze Liu, Yun Liu, Che Jiang, Kangbo Lyu, Weikang Wan, Hao Shen, Boqiang Liang, Zhoujie Fu, He Wang, and Li Yi. HOI4D: A 4D egocentric dataset for category-level human-object interaction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21013–21022, 2022. 2

[39] Zhaoyang Lv, Nicholas Charron, Pierre Moulon, Alexander Gamino, Cheng Peng, Chris Sweeney, Edward Miller, Huixuan Tang, Jeff Meissner, Jing Dong, et al. Aria everyday activities dataset. *arXiv preprint arXiv:2402.13349*, 2024. 2

[40] Juncheng Ma et al. HumanScale: Egocentric human video can outperform real-robot data for embodied pretraining. *arXiv preprint arXiv:2606.20521*, 2026. 2, 3

[41] Lingni Ma, Yuting Ye, Fangzhou Hong, Vladimir Guzov, Yifeng Jiang, Rowan Postyeni, Luis Pesqueira, Alexander Gamino, Vijay Baiyya, Hyo Jin Kim, et al. Nymeria: A massive collection of multimodal egocentric daily motion in the wild. In *European Conference on Computer Vision (ECCV)*, 2024. 2, 3

[42] Riku Murai, Eric Dexheimer, and Andrew J. Davison. MASt3R-SLAM: Real-time dense SLAM with 3D reconstruction priors. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025. 3

[43] Xiaqing Pan, Nicholas Charron, Yongqian Yang, Scott Peters, Thomas Whelan, Chen Kong, Omkar Parkhi, Richard Newcombe, and Yuheng Carl Ren. Aria digital twin: A new benchmark dataset for egocentric 3D machine perception. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 20133–20143, 2023. 2

[44] Georgios Pavlakos, Dandan Shan, Ilija Radosavovic, Angjoo Kanazawa, David Fouhey, and Jitendra Malik. Reconstructing hands in 3D with transformers. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9826–9836, 2024. 3, 4

[45] Rolandos Alexandros Potamias, Jinglei Zhang, Jiankang Deng, and Stefanos Zafeiriou. WiLoR: End-to-end 3D hand localization and reconstruction in-the-wild. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025. arXiv:2409.12259. 3, 4

[46] Ryan Punamiya et al. EgoVerse: An egocentric human dataset for robot learning from around the world. *arXiv preprint arXiv:2604.07607*, 2026. 2, 3

[47] Tong Qin, Peiliang Li, and Shaojie Shen. VINS-Mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics*, 34(4):1004–1020, 2018. 3

[48] Ri-Zhao Qiu, Shiqi Yang, Xuxin Cheng, Chaitanya Chawla, Jialong Li, Tairan He, Ge Yan, David J. Yin, Lars Paulsen Hong, Ge Yuan, et al. Humanoid policy $\sim$ human policy. *arXiv preprint arXiv:2503.13441*, 2025. 3

[49] Helge Rhodin, Christian Richardt, Dan Casas, Eldar Insafutdinov, Mohammad Shafiei, Hans-Peter Seidel, Bernt Schiele, and Christian Theobalt. EgoCap: Egocentric marker-less motion capture with two fisheye cameras. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, 35(6):162:1–162:11, 2016. 3

[50] Patrick Rim, Kevin Harris, Braden Copple, Shangchen Han, Xu Xie, Ivan Shugurov, Sizhe An, He Wen, Alex Wong, Tomas Hodan, and Kun He. Show3d: Capturing scenes of 3d hands and objects in the wild, 2026. 2

[51] Javier Romero, Dimitrios Tzionas, and Michael J. Black. Embodied hands: Modeling and capturing hands and bodies together. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, 36(6):245:1–245:17, 2017. 3, 4

[52] Yu Rong, Takaaki Shiratori, and Hanbyul Joo. FrankMocap: A monocular 3D whole-body pose estimation system via regression and integration. In *IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2021. 3

[53] Ropedia. Xperience-10M: Interactive intelligence from human xperience. https://huggingface.co/datasets/ropedia-ai/xperience-10m, 2026. 2, 3

[54] Antoni Rosinol, Marcus Abate, Yun Chang, and Luca Carlone. Kimera: An open-source library for real-time metric-semantic localization and mapping. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 1689–1696, 2020. 3

[55] Zhihao Sun et al. Ego4WAM: What matters when scaling egocentric human data for robot learning? *arXiv preprint arXiv:2609.40341*, 2026. 2

[56] Tony Tao, Mohan Kumar Srirama, Jason Jingzhou Liu, Kenneth Shaw, and Deepak Pathak. Dexwild: Dexterous human interactions for in-the-wild robot policies, 2026. 3

[57] Zachary Teed and Jia Deng. DROID-SLAM: Deep visual SLAM for monocular, stereo, and RGB-D cameras. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 16558–16569, 2021. 3

[58] Vladyslav Usenko, Nikolaus Demmel, David Schubert, Jörg Stückler, and Daniel Cremers. Visual-inertial mapping with non-linear factor recovery. *IEEE Robotics and Automation Letters*, 5(2):422–429, 2020. 3, 4

[59] Chen Wang, Haochen Shi, Weizhuo Wang, Ruohan Zhang, Li Fei-Fei, and C. Karen Liu. DexCap: Scalable and portable mocap data collection system for dexterous manipulation. In *Robotics: Science and Systems (RSS)*, 2024. 3

[60] Jianyuan Wang, Minghao Chen, Nikita Karaev, Andrea Vedaldi, Christian Rupprecht, and David Novotny. VGGT: Visual geometry grounded transformer. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025. 3

[61] Weipeng Xu, Avishek Chatterjee, Michael Zollhöfer, Helge Rhodin, Pascal Fua, Hans-Peter Seidel, and Christian Theobalt. Mo^{2}Cap^2: Real-time mobile 3D motion capture with a cap-mounted fisheye camera. *IEEE Transactions*

on Visualization and Computer Graphics, 25(5):2093–2101, 2019. 3

[62] Lixin Yang, Kailin Li, Xinyu Zhan, Fei Wu, Anran Xu, Liu Liu, and Cewu Lu. Oakink: A large-scale knowledge repository for understanding hand-object interaction, 2022. 3

[63] Lixin Yang, Jian Xu, Licheng Zhong, Xinyu Zhan, Zhicheng Wang, Kejian Wu, and Cewu Lu. POEM: Reconstructing hand in a point embedded multi-view stereo. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21108–21117, 2023. 3

[64] Lixin Yang, Licheng Zhong, Pengxiang Zhu, Xinyu Zhan, Junxiao Kong, Jian Xu, and Cewu Lu. Multi-view hand reconstruction with a point-embedded transformer, 2024. 3

[65] Zhengdi Yu, Stefanos Zafeiriou, and Tolga Birdal. Dyn-HaMR: Recovering 4D interacting hand motion from a dynamic camera. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025. 3

[66] Fan Zhang, Valentin Bazarevsky, Andrey Vakunov, Andrei Tkachenka, George Sung, Chuo-Ling Chang, and Matthias Grundmann. MediaPipe hands: On-device real-time hand tracking. *arXiv preprint arXiv:2006.10214*, 2020. 3

[67] Jinglei Zhang, Jiankang Deng, Chao Antonopoulos, Rolandos Alexandros Potamias, and Stefanos Zafeiriou. HaWoR: World-space hand motion reconstruction from egocentric videos. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025. 3

[68] Zihan Zhu, Wei Zhang, Moyang Li, Norbert Haala, Marc Pollefeys, and Daniel Barath. Vigs-slam: Visual inertial gaussian splatting slam, 2026. 3

RoboCap-10k: A Metrically Accurate Egocentric Dataset for Robot Learning

Supplementary Material

A. Additional Taxonomy Tables

	Food & beverage	Furniture & wood	Automotive	Manufacturing	Agriculture	Clothing & fashion	Construction	Household	Logistics	Retail	Crafts	Printing & packaging	Beauty & care	Cleaning & laundry	Repair	Healthcare	Laboratory	Energy	Office	Hospitality	Sports	Other / mixed	**Verticals**
India	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	21
Nigeria	●	·	●	●	●	●	●	●	·	·	●	●	●	●	●	·	·	●	·	·	·	●	13
Colombia	●	●	●	●	●	●	●	·	●	●	●	·	●	·	·	●	·	·	·	·	·	·	12
United States	·	●	●	●	·	·	●	●	·	●	●	·	·	·	·	·	·	·	·	·	·	●	7
United Kingdom	●	·	·	·	·	●	·	●	·	●	·	·	·	·	·	·	·	·	·	·	·	·	4
Mexico	·	·	●	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	●	1
Chile	·	·	●	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	1
Spain	·	·	·	·	·	·	·	·	●	·	·	·	·	·	·	·	·	·	·	·	·	·	1
Countries	4	3	6	4	3	4	4	4	3	4	4	2	3	2	2	2	1	2	1	1	1	4	

Table 10. **Geographic presence by industry vertical** across 8 countries on five continents. A dot means footage of that vertical exists in that country. The right column counts verticals per country, excluding Other / Mixed Trades; the bottom row counts countries per vertical.

	Agriculture	*Appliance repair*	*Cleaning*	*Clothing & laundry*	*Construction*	*Craft & artisan*	*Electronics*	*Food prep.*	*Furniture*	*Healthcare*	*Mechanical*	*Organization*	*Retail ops.*	*Tool use*	***Total***
Food & Beverage			4					106				4	2		116
Furniture & Woodworking				2		14			50					1	67
Automotive & Transport			13								53				66
Industrial Manufacturing					4		16					16		16	52
Agriculture & Farming	56														56
Clothing & Fashion				43		5									48
Construction & Hardware					43										43
Household / Domestic			13					3				4		1	21
Logistics & Warehousing												21			21
Retail & Consumer Goods												3	17		20
Creative Workshops & Crafts						20									20
Printing, Packaging & Design						1						2		10	13
Beauty & Personal Care						2				1					3
Cleaning, Laundry & Sanit.			3	8											11
Repair Services		1					8							2	11
Healthcare & Pharmacy										2		1			3
Laboratory / Scientific							2							2	4
Energy & Utilities					1		1							1	3
Administrative & Office												1			1
Other / Mixed Trades					1							2		2	5
Total	56	1	33	53	49	42	27	109	50	3	53	54	19	35	584

Table 11. **Manual skill groups across industry verticals**: distinct tasks per vertical and skill group, with shading proportional to count. 584 of the 619 tasks have a skill group; the remaining 35 await review. Hospitality and Sports have no labeled tasks yet and are omitted.

Vertical	Site type	Example tasks	Difficulty	Skill group
Food & Beverage	Commercial bakery	Carrying crates in bakery; Mixing and dividing dough; Moving dough loaves in tins	Medium	Food Preparation & Cooking
	Ready meal production line	Moving meal trays to machine; Portioning sauce onto pasta trays; Portioning sauce and washing trays	Medium	Food Preparation & Cooking
Furniture & Woodworking	Cane furniture workshop	Making rattan furniture; Assembling cane chair frame; Binding cane onto chair frame	Hard	Craft & Artisan Work
	Sofa upholstery workshop	Stapling fabric onto sofa frame; Stapling fabric to sofa base; Trimming foam padding for sofa	Medium	Furniture & Assembly
Automotive & Transport	Car repair garage	Draining oil under lifted car; Removing radiator and intercooler parts; Working under vehicle bonnet	Hard	Mechanical / Automotive
	Car wash / auto detailing yard	Cleaning car interior and doors; Pressure washing car wheels	Medium	Cleaning & Sanitation
Industrial Manufacturing	Electronics assembly bench	Inspecting circuit board against checklist; Soldering components on circuit board; Handling barcode label roll	Hard	Electronics & Diagnostics
	PVC scrap sorting yard	Cutting PVC pipe offcuts; Picking and cutting scrap pipes; Sorting PVC fittings into sacks	Medium	Organization & Inventory
Agriculture & Farming	Onion field	Gathering onions from field rows; Bagging harvested onions in field; Bagging onions into mesh sacks	Medium	Agriculture & Harvest
	Mango wholesale market shed	Loading mangoes into crates; Sorting and handling mangoes; Sorting and packing mangoes	Medium	Agriculture & Harvest
Clothing & Fashion	Garment stitching workshop	Sewing red fabric on machine; Sewing striped fabric on machine; Stitching white garment on machine	Hard	Clothing & Laundry
Construction & Hardware	Rooftop building site	Plastering parapet wall with trowel; Smoothing and brushing concrete surface; Tying rebar and moving materials	Hard	Construction & Building
Logistics & Warehousing	Package delivery warehouse	Sorting packages; Labeling packages; Packing packages	Easy	Organization & Inventory
Retail & Consumer Goods	Hardware store	Browsing tool and lock shelves; Picking product from shelf; Handling packaged hardware items	Easy	Retail Operations
Creative Workshops & Crafts	Home jewellery workbench	Packing jewellery into storage boxes; Sorting beaded jewellery into boxes; Assembling small jewellery pieces by hand	Medium	Craft & Artisan Work
Repair Services	Electronics repair shop	Repairing electronics; Repairing phones and laptops; Repairing electricals	Hard	Electronics & Diagnostics
Laboratory / Scientific	Clinical diagnostic lab	Operating analyser and handling samples; Pipetting samples at lab bench	Hard	Electronics & Diagnostics
Energy & Utilities	Solar site	Installing solar modules	Hard	Construction & Building

Table 12. **Representative site types and tasks**, a subset of the 365 site types. Up to three example tasks per site are taken verbatim from the catalog; difficulty and skill group are each site's typical values. The complete catalog is released with the dataset.

B. Hand Pose Gallery

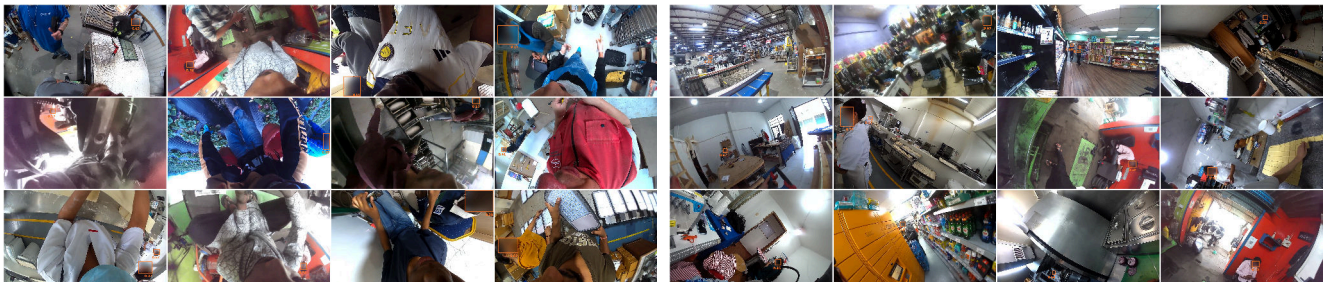


Figure 7. **Hand pose estimates across workplaces (1/2).** Thirty-six further sessions extending Figure 2, one frame each, chosen to span workplaces, tools and hand–object contact. Estimated MANO hand meshes are rendered over egocentric frames (blue: left hand, magenta: right hand).



Figure 8. **Hand pose estimates across workplaces (2/2).** Continued from Figure 7.

C. Face Blurring



(a) Eye cameras (b) Lateral cameras

Figure 9. **Additional face anonymization examples** from the eye and lateral cameras, where faces are strongly distorted or oriented.

D. Full Task Taxonomy

Table 13. **Complete site-type catalogue.** Every workplace site type in RoboCap-10k, grouped by industry vertical, with the tasks most often recorded there, its typical difficulty and its skill group. Counts are of distinct labels.

Site type	Example tasks	Difficulty	Skill group
Food & Beverage (39 site types)			
Bakery	Baking bread and pastry	Medium	Food Preparation & Cooking
Bakery and pastry	Make bread; Make pastry; Roll dough; Fill pastry; Bake bread	Medium	Food Preparation & Cooking
Bakery and pastry shop	Making baked goods; Cleaning and food preparation; Serving at front desk; Decorating pastry	Medium	Food Preparation & Cooking
Bakery cake kitchen	Icing and layering cake; Spreading icing on cake	Hard	Food Preparation & Cooking
Bakery front desk	Serve at front desk; Serving at front desk	Medium	Retail Operations
Bakery kitchen	Baking bread and pastry; Oven cooking	Medium	Food Preparation & Cooking
Biscuit packing bakery	Sealing packed biscuit trays	Medium	Food Preparation & Cooking
Bottled water plant	Handling water bottles in plant	Medium	Organization & Inventory
Bottled water warehouse	Stacking water bottles on pallets	Medium	Organization & Inventory
Butchery		Hard	Food Preparation & Cooking
Cake bakery kitchen	Decorating iced cake; Smoothing buttercream onto cake	Hard	Food Preparation & Cooking
Candy factory	Making candy	Medium	Food Preparation & Cooking
Candy manufacturing	Making candy	Medium	Food Preparation & Cooking
Candy manufacturing plant	Making candy; Packing candy; Vegetable prep	Medium	Food Preparation & Cooking
Candy production	Making candy	Medium	Food Preparation & Cooking
Catering kitchen	Catering food; Cooking fried rice	Medium	Food Preparation & Cooking
Central kitchen		Medium	Food Preparation & Cooking
Cheese sticks factory	Roll cheese sticks; Pack cheese sticks; Fry cheese sticks	Easy	Food Preparation & Cooking
Commercial bakery	Carrying crates in bakery; Mixing and dividing dough; Moving dough loaves in tins; Placing dough into bread tins; Weighing and handling bread dough	Medium	Food Preparation & Cooking
Commercial bakery kitchen	Shaping dough and trays of buns; Operating bakery equipment and handling pans; Filling pastry squares with filling; Handling dough and bakery trays; Shaping dough into loaves	Medium	Food Preparation & Cooking
Commercial food packing kitchen	Weighing and packing food mix	Medium	Food Preparation & Cooking
Commercial kitchen	Carrying bagged food product; Cleaning stainless kitchen equipment; Carrying food crates and packing; Prepping food at kitchen counter; Handling broom and chopping vegetables	Medium	Food Preparation & Cooking
Commercial kitchen bar	Making cocktails and mocktails	Medium	Food Preparation & Cooking
Commercial kitchen prep area	Weighing granola into crates	Medium	Food Preparation & Cooking
Commercial restaurant kitchen	Handling pans and plating food; Carrying and serving curry dishes; Carrying bowls and frying food; Stir-frying food in wok	Medium	Food Preparation & Cooking

Continued on next page

Site type	Example tasks	Difficulty	Skill group
Flour mill plant	Carrying flour basins to hopper; Feeding flour into mill machine; Handling flour at mill machine; Scooping flour with basin	Medium	Food Preparation & Cooking
Food packing warehouse floor	Sweeping warehouse floor	Easy	Cleaning & Sanitation
Food preparation area	Preparing food; Making empanadas	Medium	Food Preparation & Cooking
Food preparation kitchen	Making empanadas	Medium	Food Preparation & Cooking
Food production facility	Sealing lids on portion pots; Weighing food into trays	Easy	Food Preparation & Cooking
Food production kitchen	Weighing and bagging dough portions; Weighing marinated meat into trays; Filling trays with rice portions; Portioning meals into trays; Weighing and portioning meals	Medium	Food Preparation & Cooking
Kitchen	Cooking; Pack pasta and sauce pots; Pack biryani, make batter; Pack chutney and spinach; Pack rice and noodles	Medium	Food Preparation & Cooking
Kitchen store and prep room	Fetching stock and weighing ingredients	Medium	Organization & Inventory
Papad flatbread production floor	Rolling dough into flat discs	Medium	Food Preparation & Cooking
Papad making workshop	Arranging papads on drying sheets	Medium	Food Preparation & Cooking
Private chef apartment kitchen	Preparing fish and vegetables	Medium	Food Preparation & Cooking
Ready meal production line	Moving meal trays to machine; Portioning sauce onto pasta trays; Portioning sauce and washing trays; Weighing sauce onto pasta trays	Medium	Food Preparation & Cooking
Restaurant		Medium	Food Preparation & Cooking
Small food processing unit	Laying papad sheets to dry	Medium	Food Preparation & Cooking
Furniture & Woodworking (27 site types)			
Bamboo furniture workshop	Bamboo furniture work; Making bamboo crafts; Cane work	Hard	Craft & Artisan Work
Bed carpentry workshop	Bed carpentry	Hard	Furniture & Assembly
Cane furniture workshop	Making rattan furniture; Assembling cane chair frame; Binding cane onto chair frame; Binding cane onto metal frame; Shaping cane for chair frame	Hard	Craft & Artisan Work
Cane works workshop	Cane work	Hard	Craft & Artisan Work
Cane workshop	Cane work	Hard	Craft & Artisan Work
Carpentry fit-out site	Taping edge banding on plywood panel	Medium	Furniture & Assembly
Carpentry workshop	Carpentry work; Carrying and positioning timber panels; Shaping plywood table top	Hard	Furniture & Assembly
Furniture factory	Making sofas; Build bed carpentry; Upholster furniture; Manufacture furniture; Upholstery	Hard	Furniture & Assembly
Furniture manufacturing	Upholstery work; Bed carpentry	Hard	Furniture & Assembly
Furniture manufacturing workshop	Upholstery; Bed carpentry	Hard	Furniture & Assembly
Furniture workshop	Making furniture; Cutting board with jigsaw; Jigsawing curved board shapes; Stapling and spraying adhesive on panels; Cutting foam and stapling frames	Hard	Furniture & Assembly
Indoor room, view blocked			
Mattress upholstery workshop	Stapling foam and fabric to frame	Medium	Furniture & Assembly
Mattress workshop	Drilling and fitting bed base cover	Medium	Furniture & Assembly
Office room being furnished	Assembling office chair base	Medium	Furniture & Assembly
Onsite carpentry fit-out	Applying adhesive to laminate sheet	Hard	Furniture & Assembly
Open-air cane furniture workshop	Splitting and shaping cane strips; Positioning metal chair frame; Working on cane chair frames	Hard	Furniture & Assembly
Open-air carpentry workshop	Cutting wood strips with saw; Handling curved wood frame parts; Marking and cutting wood pieces; Using pneumatic stapler on frame; Marking and cutting chair frame	Hard	Furniture & Assembly
Roadside cane furniture workshop	Weaving cane onto chair frame; Preparing cane strips for chair frames	Hard	Craft & Artisan Work
Sofa and mattress workshop	Fitting foam and fabric onto furniture	Medium	Furniture & Assembly
Sofa frame workshop	Checking phone between frame work; Cutting and stapling upholstery board	Medium	Furniture & Assembly
Sofa upholstery workshop	Stapling fabric onto sofa frame; Stapling fabric to sofa base; Trimming foam padding for sofa; Fitting foam onto sofa frame; Cutting and marking board panels	Medium	Furniture & Assembly
Upholstery stitching workshop	Sewing red upholstery covers; Sewing red upholstery fabric	Hard	Clothing & Laundry
Upholstery workshop	Upholstery; Upholster furniture; Cutting and gluing fabric pieces; Fitting fabric onto sofa frame; Laying out cut upholstery fabric pieces	Hard	Furniture & Assembly
Wood workshop	Wood work	Hard	Furniture & Assembly
Wooden handicraft workshop	Sanding wooden star cutouts	Medium	Craft & Artisan Work
Woodwork shop		Hard	Furniture & Assembly

Continued on next page

Site type	Example tasks	Difficulty	Skill group
Automotive & Transport (36 site types)			
Auto repair shop	Repairing vehicles	Hard	Mechanical / Automotive
Auto wash station	Wiping car body panels	Easy	Cleaning & Sanitation
Auto workshop		Hard	Mechanical / Automotive
Bicycle repair shop	Adjusting bicycle rear derailleur; Repair bicycles	Hard	Mechanical / Automotive
Bicycle repair workshop	Handling bicycle wheel and parts; Moving and adjusting bicycles; Servicing bicycle drivetrain; Working on bicycle wheel hub; Assembling bicycle fork and handlebars	Medium	Mechanical / Automotive
Bicycle shop		Hard	Mechanical / Automotive
Bike detailing shop	Detailing bikes	Medium	Cleaning & Sanitation
Bike parts workshop	Making bike parts	Medium	Tool Use & Technical Work
Bike repair shop	Apply vinyl resin; Repairing bikes; Repair bikes	Medium	Mechanical / Automotive
Bike workshop	Repairing bikes	Hard	Mechanical / Automotive
Car detailing bay	Cleaning car door and dashboard; Wiping down car seats interior	Medium	Cleaning & Sanitation
Car detailing shop	Detailing cars	Medium	Cleaning & Sanitation
Car detailing studio	Detailing cars	Medium	Cleaning & Sanitation
Car detailing workshop	Applying polish to car body; Spraying and wiping car body	Medium	Mechanical / Automotive
Car detailing yard	Cleaning car interior seats	Medium	Mechanical / Automotive
Car interior	Cleaning car seats and dashboard	Medium	Cleaning & Sanitation
Car repair body shop	Car bodywork and painting	Hard	Mechanical / Automotive
Car repair garage	Draining oil under lifted car; Removing radiator and intercooler parts; Working under vehicle bonnet; Assembling engine parts on bench; Inspecting clutch disc	Hard	Mechanical / Automotive
Car repair shop	Repairing car; Changing tyres; Repairing car ac	Hard	Mechanical / Automotive
Car wash	Washing cars; Detailing cars; Wash cars	Easy	Cleaning & Sanitation
Car wash / auto detailing yard	Cleaning car interior and doors; Pressure washing car wheels	Medium	Cleaning & Sanitation
Car wash and detailing yard	Washing and drying car mats	Medium	Cleaning & Sanitation
Car wash yard	Pressure washing car and mats	Medium	Cleaning & Sanitation
Cycle shop		Hard	Mechanical / Automotive
Garage	Repair automobiles	Hard	Mechanical / Automotive
Motorcycle repair garage	Dismantling motorcycle frame parts	Hard	Mechanical / Automotive
Motorcycle repair shop	Repairing motorcycles; Repairing motorcycle dashboard; Dismantling and cleaning engine parts; Handling fuel tank and suspension parts; Working on motorcycle parts	Hard	Mechanical / Automotive
Motorcycle wash and service shop	Pressure washing motorcycle; Soaping and rinsing motorcycle	Medium	Mechanical / Automotive
Motors workshop		Hard	Mechanical / Automotive
Outdoor vehicle repair yard	Polishing car body panel	Medium	Mechanical / Automotive
Street bicycle repair shop	Adjusting bicycle chain and gears; Handling bicycle and tools; Servicing bicycle wheel and tyre; Working on bicycle frame and parts	Medium	Mechanical / Automotive
Street car repair shop	Repairing car headlight with tools	Hard	Mechanical / Automotive
Tyre shop		Medium	Mechanical / Automotive
Vehicle detailing	Detailing vehicles	Medium	Cleaning & Sanitation
Vehicle workshop	Repairing car; Painting car; Car paint job	Hard	Mechanical / Automotive
Workshop	Repairing car	Hard	Mechanical / Automotive
Industrial Manufacturing (40 site types)			
Assembly site	Assembly and installation	Medium	Tool Use & Technical Work
Bike parts manufacturing site	Making bike parts	Medium	Tool Use & Technical Work
Chemical plant		Medium	
Disposable plate workshop	Handling plate stacks and phone	Easy	
Electronics assembly bench	Inspecting circuit board against checklist; Soldering components on circuit board; Handling barcode label roll; Soldering components onto pcb panels	Hard	Electronics & Diagnostics
Electronics assembly factory	Handling component trays at bench; Placing circuit boards on foam trays; Soldering components on circuit board	Medium	Electronics & Diagnostics
Electronics assembly line	Assembling circuit boards on line; Fitting pcbs into panel assemblies; Assembling blue plastic components; Packing speaker units into trays; Placing speaker units into trays	Medium	Electronics & Diagnostics
Factory			
Factory bench			
Factory floor	Walking across factory floor		

Continued on next page

Site type	Example tasks	Difficulty	Skill group
Factory qc station	Filling defect inspection sheet	Medium	
Hardware factory			
Leaf plate workshop	Sorting and stacking leaf plates; Stacking leaf plates into bundles	Easy	Organization & Inventory
Metal fabrication industry		Hard	Tool Use & Technical Work
Metal fabrication workshop	Fabricating metal; Handling metal pipe sections; Handling metal rods and tools; Operating machine on shop floor; Assembling metal window frame	Hard	Tool Use & Technical Work
Metal grill fabrication workshop	Welding steel gate grill	Hard	Construction & Building
Metal tube factory	Handling metal tubes at machine	Medium	Tool Use & Technical Work
Metal tube rolling factory	Feeding steel tubes into machine	Hard	Tool Use & Technical Work
Paper plate workshop	Collecting plates from pressing machine; Counting and bundling paper plates; Passing stacked plates for packing; Stacking pressed paper plates	Easy	Organization & Inventory
Pcb assembly factory	Handling bare pcb panels; Trimming component leads on pcb	Medium	Electronics & Diagnostics
Pcb conveyor machine	Watching boards move through machine	Medium	Electronics & Diagnostics
Plastic parts production line	Handling moulded plastic parts	Easy	
Powder coating shop	Powder coating and painting	Medium	Tool Use & Technical Work
Pvc granules plant		Easy	
Pvc pipe plant		Easy	
Pvc pipe scrap yard	Sorting pvc pipe offcuts; Picking pvc pipe pieces from heap	Easy	Organization & Inventory
Pvc pipe works	Handling pvc pipes; Grading pvc pipes	Easy	Organization & Inventory
Pvc plant	Grading pvc	Easy	
Pvc plastic pipe works	Handling pvc pipes	Easy	
Pvc recycling plant	Recycling pvc; Handling paper plates	Easy	
Pvc recycling unit	Recycling pvc	Easy	
Pvc scrap sorting yard	Cutting pvc pipe offcuts; Picking and cutting scrap pipes; Sorting pvc fittings into sacks; Cutting and sorting pvc pipes; Sorting plastic pipe scrap	Medium	Organization & Inventory
Pvc works	Handling pvc	Easy	
Scrap metal sorting yard	Sorting metal scrap by hand; Sorting metal scrap into bowls	Easy	Organization & Inventory
Semiconductor plant		Medium	Electronics & Diagnostics
Spare parts factory	Manufacture spare parts; Manufacturing spare parts	Medium	Tool Use & Technical Work
Unknown		Medium	
Water bottle production	Handling water bottles	Easy	
Welding workshop	Welding metal	Hard	Tool Use & Technical Work
Window fabrication factory	Carrying vinyl frame extrusions; Feeding profiles into forming machine; Fitting seal into glass unit; Handling window frame parts	Medium	Construction & Building
Agriculture & Farming (32 site types)			
Apiary	Bee keeping	Hard	Agriculture & Harvest
Aquaculture site		Easy	Agriculture & Harvest
Crop and fish farm	Farming crops and fish	Medium	Agriculture & Harvest
Crop field	Picking and clearing field plants; Weeding and pulling plants	Medium	Agriculture & Harvest
Farm		Easy	Agriculture & Harvest
Feed and grain storage yard	Handling feed sacks and grain	Medium	Agriculture & Harvest
Field edge scrubland	Cutting grass with sickle	Medium	Agriculture & Harvest
Field under trees	Sorting harvested grass fodder	Easy	Agriculture & Harvest
Fruit and vegetable site	Organizing fruits and vegetables	Easy	Agriculture & Harvest
Grading site	Grading produce	Easy	Agriculture & Harvest
Grain grading machine shed	Scooping grain from sorting machine	Medium	Agriculture & Harvest
Green house		Easy	Agriculture & Harvest
Greenhouse farm yard	Handling irrigation line and crops	Medium	Agriculture & Harvest
Groundnut grading stall	Sorting groundnuts, handling purse	Easy	Agriculture & Harvest
Groundnut site		Easy	Agriculture & Harvest
Groundnut sorting yard	Scooping and sorting groundnuts; Sorting groundnuts in tray; Sorting groundnuts into bowls	Easy	Agriculture & Harvest
Mango wholesale market shed	Loading mangoes into crates; Sorting and handling mangoes; Sorting and packing mangoes; Wrapping mangoes in newspaper	Medium	Agriculture & Harvest
Onion field	Gathering onions from field rows; Bagging harvested onions in field; Bagging onions into mesh sacks; Trimming and sorting harvested onions; Trimming harvested onions	Medium	Agriculture & Harvest
Onion plantation	Onion planting work	Easy	Agriculture & Harvest
Open field plot	Planting seedlings in soil; Planting seedlings with hand trowel; Planting seedlings by hand; Digging soil with hand trowel; Digging soil with hand tool	Medium	Agriculture & Harvest

Continued on next page

Site type	Example tasks	Difficulty	Skill group
Outdoor groundnut sorting yard	Sorting groundnuts in tray; Picking groundnuts from tray	Easy	Agriculture & Harvest
Plant nursery	Potting plants at benches; Pruning and tending potted plants; Watering nursery plants with hose; Handling potted seedlings; Tending potted plants	Easy	Agriculture & Harvest
Plant nursery yard	Planting seedlings in soil; Watering and tending potted plants	Easy	Agriculture & Harvest
Polytunnel pepper farm	Opening seed or fertiliser packet	Medium	Agriculture & Harvest
Produce grading shed	Sieving and bagging dried produce	Medium	Agriculture & Harvest
Produce sorting shed	Sorting and bagging dried produce	Medium	Agriculture & Harvest
Red onion field	Sorting harvested onions on ground; Trimming and sorting onions; Sorting and bagging red onions	Medium	Agriculture & Harvest
Rice mill	Milling rice	Easy	Agriculture & Harvest
Strawberry plantation	Strawberry plantation work	Easy	Agriculture & Harvest
Tilled field	Digging soil with hand tool; Planting seedlings from mesh bag; Planting with hand hoe	Medium	Agriculture & Harvest
Tin-roofed shed		Easy	
Vegetable field	Picking beans from plants	Easy	Agriculture & Harvest
Clothing & Fashion (34 site types)			
Bag manufacturing factory		Medium	Clothing & Laundry
Bedroom garment sorting area	Folding clothes into storage baskets	Easy	Clothing & Laundry
Clothing and fashion			Clothing & Laundry
Clothing factory	Sewing garments; Sewing clothes	Medium	Clothing & Laundry
Clothing manufacturing	Sewing clothes; Sewing garments; Sewing clothing	Medium	Clothing & Laundry
Clothing manufacturing workshop	Sewing; Sewing garments; Sewing clothes making; Cutting fabric	Medium	Clothing & Laundry
Fabric industry	Handling fabric	Easy	
Fabric industry workshop			Clothing & Laundry
Fashion workshop	Stitching garments	Hard	Clothing & Laundry
Footwear components factory	Cutting shoe upper pieces	Medium	Craft & Artisan Work
Footwear upper cutting unit	Marking and cutting shoe panels	Medium	Craft & Artisan Work
Garment factory	Stitching striped garment panels	Hard	Clothing & Laundry
Garment factory floor	Cutting and trimming garment pieces; Handling checked fabric at table; Stitching yellow fabric at machine; Trimming and handling stitched fabric; Sewing denim garment on machine	Medium	Clothing & Laundry
Garment finishing area	Folding and stacking red garments	Medium	Clothing & Laundry
Garment sewing workshop	Sewing fabric on industrial machine	Hard	Clothing & Laundry
Garment stitching factory	Moving between sewing machines	Medium	Clothing & Laundry
Garment stitching workshop	Sewing red fabric on machine; Sewing striped fabric on machine; Stitching white garment on machine; Trimming threads on stitched shirt	Hard	Clothing & Laundry
Home garment packing room	Folding and packing garments	Easy	Clothing & Laundry
Home garment sorting room	Folding garments on bed	Easy	Clothing & Laundry
Home garment workshop	Folding and stacking finished garments	Medium	Clothing & Laundry
Home tailoring workroom	Ironing pressing fabric pieces	Medium	Clothing & Laundry
Seat cover workshop	Making seat covers	Hard	Craft & Artisan Work
Sewing factory	Sew garments; Rivet	Medium	Clothing & Laundry
Sewing manufacturing workshop	Sewing; Cutting and sewing	Medium	Clothing & Laundry
Sewing workshop	Sew garments; Sewing; Sewing fabric on industrial machine; Stitching printed fabric cover	Medium	Clothing & Laundry
Shoemaking workshop	Make shoes	Hard	Craft & Artisan Work
Small tailoring workshop	Sorting and handling cut garments	Medium	Clothing & Laundry
Stitching workshop	Stitching	Hard	Clothing & Laundry
Tailor shop		Hard	Clothing & Laundry
Tailoring shop	Tailoring garments; Stitching fabric at sewing machine; Threading machine and handling fabric; Setting up sewing machine; Stitching fabric on sewing machine	Hard	Clothing & Laundry
Tailoring workshop	Stitching fabric on sewing machine; Feeding fabric through sewing machine; Pressing and finishing garments; Sewing embellished garment on machine	Hard	Clothing & Laundry
Tailors shop		Hard	Clothing & Laundry
Textile workplace			Clothing & Laundry
Weaving workshop	Weaving on loom	Hard	Craft & Artisan Work

Continued on next page

Site type	Example tasks	Difficulty	Skill group
Construction & Hardware (34 site types)			
Brick building construction site	Handling wiring and hammering timber	Hard	Construction & Building
Brick works		Easy	Construction & Building
Building construction site	Positioning formwork panels and rebar	Hard	Construction & Building
Building interior site	Fitting metal frame and wiring	Hard	Construction & Building
Building interior under finishing	Painting and finishing walls	Medium	Construction & Building
Building site	Laying bricks; Bricklaying and iron bending	Hard	Construction & Building
Building site interior	Carrying bamboo scaffolding poles; Handling timber and concrete materials; Mixing and scooping cement	Medium	Construction & Building
Building site rooftop	Mixing and carrying mortar	Medium	Construction & Building
Building site staircase	Plastering concrete stairs with trowel	Hard	Construction & Building
Building under construction	Carrying cement pan up stairs	Medium	Construction & Building
Concrete building under construction	Chiselling holes in concrete ceiling; Running electrical conduit wiring	Hard	Construction & Building
Construction hardware			Construction & Building
Construction hardware site		Hard	Construction & Building
Construction site	Construction work; Iron rebar operations; Bending iron	Hard	Construction & Building
Construction site yard	Bending and tying steel reinforcement bars	Hard	Construction & Building
Dark unknown		Easy	
Dark unlit interior	Diy workshop work		
Granite works		Hard	Construction & Building
High-rise construction site	Tying rebar cage	Hard	Construction & Building
Home construction site		Hard	Construction & Building
Interior fit-out site	Applying filler to door frame; Sanding and finishing interior walls	Medium	Construction & Building
Outdoor fencing site	Installing fence posts on slope	Medium	Construction & Building
Painting site	Painting	Medium	Construction & Building
Roadside shop canopy build	Fitting metal roof canopy frame	Hard	Construction & Building
Roadside shopfront build	Drilling into shopfront structure	Hard	Construction & Building
Rooftop building site	Plastering parapet wall with trowel; Smoothing and brushing concrete surface; Tying rebar and moving materials	Hard	Construction & Building
Rooftop metal railing install	Cutting and fitting steel railing	Hard	Construction & Building
Tiling site	Lay tiles	Hard	Construction & Building
Unfinished brick building	Diy workshop work		
Unfinished building interior	Sweeping construction floor debris	Easy	Construction & Building
Window and glass factory	Handling glass panes on racks	Medium	Construction & Building
Window fabrication workshop	Tapping glass unit with mallet	Medium	Construction & Building
Window frame workshop	Fitting rubber seal into frame	Medium	Construction & Building
Window manufacturing plant	Applying sealant to window frame; Carrying window frames to rack; Fitting glass into window frame; Fitting sash onto glass unit; Loading frames into welding machine	Medium	Construction & Building
Household / Domestic (15 site types)			
Home	Home tasks; Organizing household items	Easy	Cleaning & Sanitation
Home balcony and bathroom	Diy workshop work		
Home bathroom	Scrubbing bathroom floor with detergent	Easy	Cleaning & Sanitation
Home corridor and rooms	Sweeping floor with broom	Easy	Cleaning & Sanitation
Home dining kitchen area	Handling small bottle at table	Easy	
Home dining table workspace	Sorting power tools on table	Medium	Organization & Inventory
Home kitchen	Washing dishes at sink; Wiping counters and stovetop; Cooking chicken breasts; Preparing food at stove; Peeling potatoes and rolling roti	Easy	Cleaning & Sanitation
Home kitchen and dining room	Handling spray cans and tools	Easy	Organization & Inventory
Home kitchen and dining table	Eating and using tablet	Easy	
Home kitchen and table	Handling tools and household items	Easy	Organization & Inventory
Home kitchen dining area	Tidying table using tablet	Easy	
Home living room carpet	Picking up items from floor	Easy	Cleaning & Sanitation
Home table tool bench	Handling hand tools at table	Medium	Tool Use & Technical Work
Household	Cleaning the house; Mopping floors; Doing laundry; Cleaning surfaces; Sweeping floors	Easy	Cleaning & Sanitation
Private home	Mopping floors; Cleaning house; General cleaning; Household chores and repair	Easy	Cleaning & Sanitation

Continued on next page

Site type	Example tasks	Difficulty	Skill group
Logistics & Warehousing (9 site types)			
Delivery route	Drive deliveries	Easy	
Handicraft packing warehouse	Foam wrapping wooden handles; Packing goods into cartons	Easy	Organization & Inventory
Office packing area	Unwrapping and sorting packed devices; Unpacking devices from bags	Easy	Organization & Inventory
Package delivery warehouse	Sorting packages; Labeling packages; Packing packages; Label inventory; Label sort dispatch inventory	Easy	Organization & Inventory
Packing warehouse	Folding cardboard boxes for packing; Sorting metal parts on bench	Easy	Organization & Inventory
Parcel sorting warehouse	Sorting packets on sorting table; Sorting parcels into crates	Easy	Organization & Inventory
Stock room	Manage inventory and stock	Easy	Organization & Inventory
Warehouse		Easy	Organization & Inventory
Warehouse pick and pack	Packing parts into cartons	Medium	Organization & Inventory
Retail & Consumer Goods (20 site types)			
Big box store	Picking boxed items from shelves	Easy	Retail Operations
Clothes store	Organizing clothes store	Easy	Retail Operations
Convenience grocery store	Facing jars on shelves	Easy	Organization & Inventory
Convenience shop	Handling products on shelves; Picking boxed products from shelves	Easy	Retail Operations
Cycle market		Easy	Retail Operations
Discount variety store	Reaching for stationery items	Easy	Retail Operations
General variety store	Handling packaged goods on shelves	Easy	Retail Operations
Grocery shop aisle	Picking products from shelves	Easy	Retail Operations
Grocery store	Taking bottles from shelf	Easy	Retail Operations
Grocery store aisle	Picking products from shelves	Easy	Retail Operations
Grocery supermarket aisle	Picking jars from grocery shelves	Easy	Retail Operations
Hardware store	Browsing tool and lock shelves; Picking product from shelf; Handling packaged hardware items	Easy	Retail Operations
Indian grocery shop	Inspecting packaged food on shelves	Easy	Retail Operations
Jewellery shop		Medium	Retail Operations
Pet store	Organizing pet store; Organising store stock; Organizing store stock	Easy	Retail Operations
Pet supply store	Walking aisles checking stock	Easy	Retail Operations
Provision store		Easy	Retail Operations
Retail		Easy	Retail Operations
Shoe store	Handling shoes on display racks	Easy	Retail Operations
Stationery store aisle	Checking shelf stock on phone	Medium	Organization & Inventory
Creative Workshops & Crafts (13 site types)			
Beadwork craft workshop	Handling bead packets on table; Sorting beads into packets	Medium	Craft & Artisan Work
Craft workspace	Crocheting	Hard	Craft & Artisan Work
Creative workshop			
Florist	Arranging flowers	Medium	Craft & Artisan Work
Flower shop		Medium	Craft & Artisan Work
Handicraft workshop floor	Assembling wooden craft discs; Handling packing sack; Assembling wrapped decorative craft items	Medium	Craft & Artisan Work
Home dining table	Placing beads on diamond painting	Easy	Craft & Artisan Work
Home jewellery workbench	Packing jewellery into storage boxes; Sorting beaded jewellery into boxes; Assembling small jewellery pieces by hand; Filing and shaping jewellery parts	Medium	Craft & Artisan Work
Home living room table	Assembling laser-cut model parts; Assembling miniature model kit	Medium	Craft & Artisan Work
Lamp making workshop	Make oil lamps; Make cotton lamp wicks	Hard	Craft & Artisan Work
Weaving workshop	Weaving on loom	Hard	Craft & Artisan Work
Wooden handicraft workshop	Sanding wooden craft boards; Sanding wooden craft pieces	Medium	Craft & Artisan Work
Wooden star craft workshop	Sanding wooden star on buffer; Sanding wooden stars on buffer	Medium	Craft & Artisan Work
Printing, Packaging & Design (14 site types)			
Corrugated board factory	Feeding board sheets through machine; Marking and stacking board sheets	Medium	Organization & Inventory
Diary production	Making diaries	Easy	
Handicraft packing workshop	Wrapping items in foam	Easy	Organization & Inventory
Offset printing press shop	Handling printed sheets and press parts; Walking around press floor	Easy	Tool Use & Technical Work
Offset printing press workshop	Handling printed sheets at press; Inspecting printed sheets and feeding press; Loading paper and checking print; Operating printing press	Hard	Tool Use & Technical Work
Packaging materials workshop	Handling packaging sheet on bench	Easy	

Continued on next page

Site type	Example tasks	Difficulty	Skill group
Paper plate making workshop	Pressing and stacking paper plates	Medium	Craft & Artisan Work
Paper plate unit	Making paper plates	Easy	
Print and fabrication shop	Printing and fabrication	Medium	Tool Use & Technical Work
Print shop	Printing; Handling printed paper stacks	Medium	Tool Use & Technical Work
Printing press		Medium	Tool Use & Technical Work
Printing press workshop	Operating printing press machine	Hard	Tool Use & Technical Work
Stationery production	Produce stationery	Easy	
Stationery workshop			
Beauty & Personal Care (13 site types)			
Barber shop	Barbing hair; Cutting hair	Hard	
Barbershop	Cutting hair	Hard	
Beauty care salon		Hard	
Beauty parlour		Hard	
Beauty salon		Hard	
Hair salon	Braiding hair; Braiding and wig revamp; Revamp wigs and locks; Revamp wigs; Cutting hair	Hard	Craft & Artisan Work
Nail bar	Nail care; Doing nails	Hard	
Nail care salon	Manicure and pedicure	Hard	
Nail salon		Hard	
Personal care setting	Personal care	Medium	
Pet shop	Grooming pets	Medium	
Pet store	Grooming pets; Grooming pets, cat surgery	Medium	Healthcare & Medical Tasks
Salon		Hard	
Cleaning, Laundry & Sanitation (13 site types)			
Cleaning site		Easy	Cleaning & Sanitation
Commercial premises	Commercial cleaning	Easy	Cleaning & Sanitation
Deep cleaning job	Deep cleaning	Easy	Cleaning & Sanitation
Deep cleaning service	Deep cleaning	Easy	Cleaning & Sanitation
Home deep cleaning	Deep cleaning home	Easy	Cleaning & Sanitation
Home ironing room	Ironing white linen on board	Medium	Clothing & Laundry
Home laundry room	Folding and ironing clothes	Medium	Clothing & Laundry
Ironing business	Ironing clothes	Easy	Clothing & Laundry
Laundry	Laundry work; Doing laundry	Medium	Clothing & Laundry
Laundry service	Doing laundry	Medium	Clothing & Laundry
Outdoor communal laundry yard	Hand washing clothes in basins	Medium	Clothing & Laundry
Outdoor hand laundry yard	Washing clothes in basins	Medium	Clothing & Laundry
Outdoor washing area	Scrubbing clothes in basin	Medium	Clothing & Laundry
Repair Services (9 site types)			
Electrical works shop	Electrical works	Hard	Tool Use & Technical Work
Electronics repair shop	Repairing electronics; Repairing phones and laptops; Repairing electricals; Repairing laptops	Hard	Electronics & Diagnostics
Mobile phone shop	Assembling disassembling phones	Hard	Electronics & Diagnostics
Mobile repair shop	Repairing mobile phones	Hard	Electronics & Diagnostics
Phone repair shop	Repair phones	Hard	Electronics & Diagnostics
Plumbing job	Plumbing work	Hard	Tool Use & Technical Work
Plumbing job site	Plumbing work	Hard	Tool Use & Technical Work
Plumbing service	Plumbing work	Hard	Tool Use & Technical Work
Repair shop	Repair phones and laptops; Repairing items	Hard	Electronics & Diagnostics
Healthcare & Pharmacy (7 site types)			
Hospital		Hard	Healthcare & Medical Tasks
Hospital medical store	Handling medicine stock at desk	Medium	Organization & Inventory
Hospital medical store counter	Entering records on computer	Medium	
Hospital operating theatre	Monitoring anaesthesia equipment	Hard	Healthcare & Medical Tasks
Indoor wall			
Pharmacy		Medium	Retail Operations
Veterinary clinic	Cat surgery	Hard	Healthcare & Medical Tasks
Laboratory / Scientific (4 site types)			

Continued on next page

Site type	Example tasks	Difficulty	Skill group
Clinical diagnostic lab	Operating analyser and handling samples; Pipetting samples at lab bench	Hard	Electronics & Diagnostics
Diagnostics centre		Hard	Healthcare & Medical Tasks
Water quality lab	Using ph and tds meter kit	Hard	Electronics & Diagnostics
Water testing laboratory	Labelling sample dishes at bench	Hard	Tool Use & Technical Work
Energy & Utilities (3 site types)			
Electrical installation site	Installing electrics	Hard	Tool Use & Technical Work
Solar site	Installing solar modules	Hard	Construction & Building
Switchyard site	Metering switchyard	Hard	Electronics & Diagnostics
Administrative & Office (3 site types)			
Office			
Office meeting room	Walking through office carrying items	Easy	
Open plan office	Carrying and unpacking boxed items	Easy	Organization & Inventory
Hospitality & Lodging (3 site types)			
Hotel		Easy	Cleaning & Sanitation
Hotel guest rooms		Easy	Cleaning & Sanitation
Hotel rooms		Easy	Cleaning & Sanitation
Sports & Recreation (2 site types)			
Gym			
Sports goods shop		Easy	Retail Operations
Other / mixed trades (5 site types)			
Car interior and hotel lobby	Driving a car	Medium	
Installation site	Install pop	Hard	Construction & Building
Material sorting site	Sort materials	Easy	Organization & Inventory
Unspecified site	Assembly and installation; Organizing and decluttering	Easy	Organization & Inventory
Workshop	Diy workshop work; Doing diy work	Medium	Tool Use & Technical Work