Sensor Preprocessing and State Estimation for Multi-Modal Robotics

Vision-based tactile sensors, LiDAR-inertial odometry, and factor graph optimization have converged to enable robust state estimation for mobile manipulation systems. For an **8-DOF arm on a mobile base** with encoders (10-50 kHz), torque sensors (1-3 kHz), tactile sensors (30-60 Hz), IMU (100 Hz), cameras (30 FPS), and LiDAR (10 Hz), the optimal architecture combines **factor graph optimization** for multi-sensor fusion with specialized preprocessing pipelines for each modality. Recent advances in learned tactile representations (Sparsh, 2024) and lightweight LIO systems (FAST-LIO2, Faster-LIO) enable real-time performance on embedded platforms like Jetson Orin.

IMU preprocessing fundamentals shape downstream fusion accuracy

IMU preprocessing establishes the foundation for all inertial-aided state estimation. Three complementary filtering approaches dominate practice:

Complementary filter exploits frequency-domain sensor characteristics—gyroscopes accurate at high frequencies, accelerometers at low frequencies— PubMed Central via the simple fusion $\theta\_est = \alpha(\theta\_est\_prev + \omega\cdot\Delta t) + (1-\alpha)\cdot\theta\_accel$, where $\alpha \approx 0.96$-0.98 for 100 Hz IMU. **Madgwick filter** (2010) uses gradient descent to minimize orientation error DeepWiki with gain $\beta \approx 0.033$, achieving higher accuracy but requiring more computation. **Mahony filter** (IEEE TAC 2008) operates directly on SO(3) with explicit PI correction, providing online gyro bias estimation suitable for hardware implementation. Readthedocs

IMU preintegration (Forster et al., TRO 2016) revolutionized visual-inertial systems by summarizing hundreds of IMU measurements between keyframes into single relative motion constraints: Robotics: Science and Systems

- $\Delta R\_ij = \prod Exp((\tilde{\omega}\_k - b\_g)\Delta t)$ for rotation
- $\Delta v\_ij = \Sigma\ \Delta R\_ik(\tilde{a}\_k - b\_a)\Delta t$ for velocity
- $\Delta p\_ij = \Sigma\ [\Delta v\_ik\cdot\Delta t + \frac{1}{2}\Delta R\_ik(\tilde{a}\_k - b\_a)\Delta t^2]$ for position

First-order Jacobians enable **bias correction without reintegration**, critical for real-time optimization. GTSAM 4.0 provides the standard open-source implementation. uzh Robotics: Science and Systems

Allan Variance characterization (15-24 hour stationary recordings) identifies noise parameters from log-log slope: angle random walk (slope -0.5), bias instability (minimum point ÷ 0.664), and rate random walk (slope +0.5). GitHub Typical MEMS values: ARW 0.1-0.5 °/√hr, bias instability 1-10 °/hr.

Proprioceptive sensing requires careful derivative estimation

High-frequency encoder data (10-50 kHz) demands robust velocity estimation. The **MT-method** combines position counting and period measurement: NCBI $v = \Delta N\cdot(2\pi/PPR)/(N\_clk\cdot T\_clk)$, providing accuracy across wide speed ranges. For smooth derivatives, **Savitzky-Golay filters** perform least-squares polynomial fitting, Stack Exchange while **Levant's super-twisting differentiator** provides finite-time convergence with bounded noise amplification through sliding mode theory.

Generalized Momentum Observer (De Luca & Mattone, ICRA 2005) estimates external torques without noisy acceleration measurement. For robot dynamics $M(q)\ddot{q} + C(q,\dot{q})\dot{q} + g(q) = \tau + \tau_{ext}$, the observer exploits the skew-symmetric property of $\dot{M} - 2C$:

$$r(t) = K_I\int[\tau - C^T\dot{q} - g - r]d\tau + p(0) - p(t)$$

where $\hat{\tau}_{ext} = r(t)$ provides first-order filtered external torque. Recent improvements include **Super-Twisting Momentum Observer** (Long et al., JIRS 2023) with sigmoid + PI structure and adaptive gains, ResearchGate and **composite observers** combining GMO with extended state observers to reduce initial peaking (Ibari et al., AIMS 2024).

LiDAR preprocessing pipelines extract geometric structure

Point cloud preprocessing follows a systematic pipeline. **Voxel grid filtering** (PCL: `setLeafSize(0.1-0.5m)`) reduces density while preserving structure. Chambbj **Statistical outlier removal** eliminates noise via Gaussian distance distribution analysis (typically `setMeanK(50)`, `setStddevMulThresh(1.0)`). Readthedocs **Ground segmentation** options include RANSAC plane fitting, **Cloth Simulation Filter** (Zhang et al., Remote Sensing 2016) with intuitive physical modeling, CloudCompare and learning-based approaches like Patchwork++.

Feature extraction follows two paradigms:

LOAM-style features (Zhang & Singh, RSS 2014) compute smoothness $c = (1/|S|)\times||\Sigma(p_j - p_i)||$ to classify edge points (high c) and planar points (low c), with maximum 2 edge and 4 planar points per sub-region. LearnOpenCV **KISS-ICP** (Vizzo et al., RA-L 2023) eliminates explicit feature extraction entirely, using adaptive threshold point-to-point ICP with Welsch robust kernel—ranking 2nd among open-source systems on KITTI with minimal tuning. arXiv

Scan matching algorithms span point-to-point ICP (simple, noise-sensitive), point-to-plane ICP (better for structured environments), **GICP** (Segal et al., RSS 2009) modeling surfaces as Gaussians (~0.85m ATE vs ~2.7m for standard ICP on KITTI), and **NDT** (voxelized Gaussian PDFs with larger convergence basins).

LIO systems achieve centimeter-level accuracy at 100+ Hz

Three LiDAR-Inertial Odometry architectures dominate:

System	Backend	Map Structure	Speed	Loop Closure
LIO-SAM (Shan et al., IROS 2020)	Factor graph (GTSAM iSAM2)	Keyframe point clouds	10-20 Hz	Yes
FAST-LIO2 (Xu et al., T-RO 2022)	Iterated EKF	ikd-Tree	Up to 100 Hz	No
Faster-LIO (Bai et al., RA-L 2022)	Iterated EKF	iVox hash map	200-2000 Hz	No

LIO-SAM employs four factor types (IMU preintegration, LiDAR odometry, GPS, loop closure) with LOAM-style features and sliding window keyframes. ResearchGate Requires 9-axis IMU at ≥200 Hz recommended. GitHub

FAST-LIO2 performs **direct point registration** without feature extraction, using an incremental k-d tree (ikd-Tree) with parallel rebuilding and a novel Kalman gain formula reducing complexity from O(measurement_dim) to O(state_dim). Handles **1000 deg/s rotation** and runs on ARM processors (Jetson TX2, Raspberry Pi 4B). ResearchGate

Faster-LIO replaces ikd-Tree with **iVox** (incremental voxels via sparse hash map), achieving **72% faster search** than ikd-Tree and **97% faster than k-d tree** through parallel k-NN queries. ResearchGate ResearchGate

Recent advances include **KISS-ICP** (parameter-free adaptive ICP), IEEE Xplore **DLIO** (ICRA 2023, continuous-time correction, 20% more efficient), **Point-LIO** (point-by-point sub-frame registration), and **LIO-GVM** (RA-L 2024, Gaussian voxel maps with variance-based outlier rejection).

Visual features span classical to learned approaches

ORB features combine FAST detection with rotated BRIEF descriptors PLOS via intensity centroid orientation. Multi-scale pyramids (8 levels, 1.2× factor) provide scale invariance. OpenCV Binary descriptors enable Hamming distance matching, achieving **100× speedup over SIFT** PLOS at ~50% of SLAM CPU time.

SuperPoint (DeTone et al., CVPR 2018 Workshop) uses self-supervised learning on synthetic shapes followed by homographic adaptation on real images. TheCVF The VGG-style encoder produces 256-dimensional float descriptors with superior repeatability under viewpoint/illumination changes. **SuperPoint-SLAM3** (Syed et al., arXiv 2025) reduces KITTI translational error from **4.15% to 0.34%**. arXiv

Learned matchers dramatically improve correspondence quality:

- **SuperGlue** (Sarlin et al., CVPR 2020): Graph neural network with optimal transport via Sinkhorn algorithm GitHub (~70ms inference)
- **LightGlue** (Lindenberger et al., ICCV 2023): Adaptive depth/width with early exit, arXiv **4-10× faster** than SuperGlue at 150 FPS @ 1024 keypoints GitHub

- **LoFTR** (Sun et al., CVPR 2021): Detector-free dense matching via Transformer attention, handles low-texture scenes

Stereo matching via Semi-Global Matching (SGM) aggregates costs along 8-16 directions with smoothness penalties P1 (small jumps) and P2 (large jumps, adaptive to gradient). Real-time: **42 FPS @ 640×480** on Tegra X1. ScienceDirect Learned alternatives (RAFT-Stereo, CreStereo) achieve superior accuracy with GPU.

VIO systems offer accuracy-computation tradeoffs

MSCKF (Mourikis & Roumeliotis, ICRA 2007) maintains a sliding window of camera poses via stochastic cloning. Feature observations create multi-view constraints marginalized through **null-space projection**, achieving $O(N^2)$ complexity for N camera clones. First-Estimates Jacobian (FEJ) maintains observability consistency.

VINS-Mono/Fusion (Qin et al., T-RO 2018) performs tightly-coupled optimization with IMU preintegration factors, visual reprojection factors, and 4-DOF pose graph loop closure. Critical initialization aligns visual scale with IMU through gyroscope bias, gravity direction, and velocity estimation. EuRoC performance: **0.05-0.15m ATE**.

ORB-SLAM3 (Campos et al., T-RO 2021) introduces the **Atlas multi-map system** surviving tracking loss with seamless map merging. Novel two-stage IMU initialization separates visual-only and inertial-only MAP estimation. Stereo-inertial achieves **3.5 cm average ATE on EuRoC**, 2.6× more accurate than VINS-Mono.

OpenVINS (Geneva et al., ICRA 2020) provides modular research-oriented MSCKF with online calibration of camera intrinsics, IMU-camera extrinsics, and time offset. Native ROS 2 support from v2.7+.

System	Method	EuRoC ATE	Computation	ROS 2
MSCKF	Filter	10-20 cm	~10 ms	Via OpenVINS
VINS-Mono	Optimization	5-15 cm	~50 ms	Community
ORB-SLAM3	Optimization	3.5 cm (stereo-inertial)	~100 ms	Community
OpenVINS	Filter	5-15 cm	~15 ms	Native

Recent innovations include **RD-VIO** (TVCG 2024) with IMU-PARSAC for dynamic environments, **PO-MSCKF** (arXiv 2024) eliminating null-space projection via pose-only theory, and **SuperVINS** (arXiv 2024) integrating SuperPoint + LightGlue into VINS-Fusion.

Vision-based tactile sensors enable sub-millimeter contact perception

GelSight sensors use transparent elastomer with reflective coating, reconstructing 3D contact geometry via photometric stereo from RGB LED illumination. MDPI PubMed Central Poisson equation solving with Discrete Sine Transform produces **~20-30 µm spatial resolution**. Marker tracking reveals force patterns: radial compression (normal force), directional displacement (shear), rotational pattern (torque).

DIGIT (Lambeta et al., RA-L 2020) provides compact 20×27mm form factor with USB connectivity and open-source design. The **TACTO simulator** enables sim-to-real transfer for learned policies.

DIGIT360 (Lambeta et al., arXiv 2024) achieves hemispherical omnidirectional sensing with **~8.3 million taxels**, Projectreylo hyperfisheye optics, and 18+ sensing modalities including vision, vibration, temperature, and chemical detection. Projectreylo On-device NPU enables real-time processing via USB-3.1 Type-C. Projectreylo ROS 2 driver available; commercial availability planned for 2025 through GelSight Inc. Meta

AnySkin (Bhirangi et al., arXiv 2024) uses magnetic tactile sensing with 5 magnetometers detecting field distortions from embedded iron particles. Unite.AI Key advantage: **12-second replacement** with only 13% performance drop across instances (vs. 43% for ReSkin). Cost: ~$10 at scale. arXiv

Tactile preprocessing encompasses:

- **Force estimation**: U-Net predicting force distributions from RGB images arXiv achieves **0.54N normal, 0.26-0.33N shear RMSE** Wiley Online Library (FEATS, 2024)
- **Slip detection**: Entropy-based marker displacement analysis reaches **95.61% accuracy** without prior object knowledge (ICRA 2023)
- **Contact geometry**: Photometric stereo pipeline with marching cubes mesh reconstruction

Sparsh (Higuera et al., CoRL 2024) provides the first tactile foundation model, pre-trained on **460K+ tactile images** via DINO/I-JEPA self-supervised learning. Sparsh-ssl Achieves **95.1% improvement** over task-specific models on TacBench tasks (force estimation, slip detection, pose estimation, grasp stability, textile recognition, dexterous manipulation). github

Sensor fusion architectures balance accuracy and computation

Extended Kalman Filter (EKF) linearizes via Taylor expansion with $O(n^3)$ complexity dominated by matrix inversion. **Unscented Kalman Filter (UKF)** uses deterministic sigma point sampling (2n+1 points) to capture second-order statistics, providing **10-30% accuracy improvement** in highly nonlinear conditions at ~3× computational cost.

Factor graph optimization via GTSAM with **iSAM2 incremental solver** offers native multi-sensor support through specialized factors: MathWorks

- **ImuFactor**: On-manifold preintegration
- **BetweenFactor**: Odometry constraints
- **GenericProjectionFactor**: Visual reprojection
- **GPSFactor**: Global position constraints

iSAM2 uses Bayes tree representation for O(n) incremental updates with automatic relinearization (RelinearizeThreshold: 0.01-0.1). **Ceres Solver** provides automatic differentiation via Jets (~1000× faster than numeric differentiation) with manifold support for SO(3)/SE(3).

Aspect	EKF	UKF	Factor Graphs
Accuracy	Moderate	High (2nd order)	Highest (batch)
Update time	<1 ms	2-3 ms	5-10 ms (iSAM2)
Loop closure	Limited	Limited	Native
Multi-sensor	Sequential	Sequential	Native

Multi-rate fusion runs at highest sensor rate (IMU at 100+ Hz) with selective update steps when lower-rate measurements arrive. Factor graphs handle asynchronous sensors through explicit timestamped constraints and IMU preintegration between keyframes.

Implementation requires careful noise characterization and synchronization

Sensor noise models (typical values):

Sensor	Key Parameters
Encoders	Resolution: 0.001-0.01 rad, density: 1e-4 rad/√Hz
Torque	Resolution: 0.1-1% FS, density: 0.01-0.1 Nm/√Hz
Tactile	Force: 0.1-0.5 N σ, position: 1-3 mm σ
IMU (consumer MEMS)	ARW: 0.3-1.0 °/√hr, bias: 10-50 °/hr
IMU (industrial)	ARW: 0.01-0.1 °/√hr, bias: 0.1-5 °/hr
Camera features	0.5-2.0 pixels σ
Stereo depth	$\sigma_z = 0.01\text{-}0.05 \times z^2$ (quadratic)
LiDAR range	2-5 cm σ

Time synchronization via IEEE 1588 PTP achieves **<100 ns** with hardware timestamping, **~100 µs-1 ms** with software. ROS 2 integration uses `message_filters::ApproximateTimeSynchronizer` for soft sync and `tf2_ros::MessageFilter` to wait for transforms before callbacks.

ROS 2 patterns include `robot_localization` package for EKF-based fusion, lifecycle nodes for managed sensor initialization, and tf2 for transform management across sensor frames.

Jetson AGX Orin deployment (275 TOPS INT8, 2048 CUDA cores, 32/64 GB LPDDR5) achieves:

- LIO-SAM: 50+ Hz
- ORB-SLAM3: 30+ Hz
- Object detection: 60+ FPS

TensorRT optimization (`trtexec --onnx=model.onnx --saveEngine=model.trt --fp16`) provides ~2× FP16 speedup; INT8 quantization achieves ~4× with calibration.

State-of-the-art research papers (2022-2025)

VIO/LIO advances:

- KISS-ICP (Vizzo et al., RA-L 2023): arxiv.org/abs/2209.15397
- DLIO (Chen et al., ICRA 2023): arxiv.org/abs/2203.03749 — continuous-time correction arXiv
- Point-LIO (He et al., Advanced Intelligent Systems 2023): Point-by-point sub-frame registration
- LIO-GVM (Ji et al., RA-L 2024): Gaussian voxel maps
- SuperPoint-SLAM3 (Syed et al., arXiv 2025): arxiv.org/abs/2506.13089
- RD-VIO (Li et al., TVCG 2024): arxiv.org/abs/2310.15072 — dynamic environment handling
- PO-MSCKF (Du et al., arXiv 2024): arxiv.org/abs/2407.01888

Tactile perception:

- Sparsh (Higuera et al., CoRL 2024): Sparsh-ssl arxiv.org/abs/2410.24090 — tactile foundation model
- DIGIT360 (Lambeta et al., arXiv 2024): GitHub arxiv.org/abs/2411.02479 — omnidirectional multimodal fingertip
- AnySkin (Bhirangi et al., arXiv 2024): arxiv.org/abs/2409.08276 — plug-and-play magnetic tactile
- 3D-ViTac (arXiv 2024): arxiv.org/abs/2410.24091 — point cloud tactile representations
- TacDiffusion (arXiv 2024): arxiv.org/abs/2409.11047 — force-domain diffusion policy

Multi-sensor fusion:

- FT-LVIO (Zhang et al., IET 2023): Fully tightly-coupled LiDAR-Visual-Inertial
- OKVIS2-X (arXiv 2024): arxiv.org/abs/2510.04612 — modular multi-sensor with LiDAR/GNSS
- UKF-Based Joint-Torque Fusion (arXiv 2024): arxiv.org/abs/2402.18380
- Learned Selective Sensor Fusion (Chen et al., IEEE TNNLS 2025): Interpretable attention mechanisms

Factor graph optimization:

- Multi-Momentum Observer Contact Estimation (arXiv 2024): arxiv.org/abs/2412.03462
- Swarm-LIO2 (T-RO 2024): Decentralized LIO for robot swarms
- SLAM2REF (Construction Robotics 2024): Multi-session anchoring with reference maps

Recommended architecture for the 8-DOF mobile manipulation platform

Preprocessing layer:

- IMU: Madgwick filter (β=0.033) + preintegration for factor graph
- Encoders: MT-method velocity estimation + Savitzky-Golay smoothing
- Torque: Generalized Momentum Observer with K_I tuned for 10-50 Hz bandwidth
- LiDAR: Voxel downsampling (0.2m) → Patchwork++ ground segmentation → direct registration
- Camera: SuperPoint + LightGlue (GPU) or ORB (CPU fallback)
- Tactile: Sparsh encoder for learned representations, Sparsh-ssl U-Net for force estimation arXiv

State estimation layer: Factor graph (GTSAM iSAM2) with:

- IMU preintegration factors (100 Hz → keyframe rate)
- LiDAR point-to-plane factors (10 Hz)
- Visual reprojection factors (30 Hz)
- Encoder odometry factors (high-rate, marginalized)
- Torque residual factors for contact detection
- Tactile factors for manipulation contact constraints

Implementation:

- Primary compute: Jetson AGX Orin with TensorRT optimization
- ROS 2 Jazzy with lifecycle nodes and message_filters synchronization
- Hardware PTP where available; software sync via tf2_ros::MessageFilter
- Latency budget: <100 ms total pipeline (IMU <1 ms, LiDAR <50 ms, visual <30 ms, fusion <20 ms)

This architecture balances state-of-the-art accuracy with real-time performance across the **10-50 kHz to 10 Hz** sensor rate spectrum, enabling robust mobile manipulation in unstructured environments.