Proceedings of the 4th International Conference on Applied Physics and Mathematical Modeling
DOI: 10.54254/2753-8818/2026.37905

Multi-sensor Calibration Technology for Intelligent Connected Vehicles

Yurong Wu

Chongqing Yucai Secondary School, Chongqing, China
WuYurong1224@hotmail.com

Abstract. Multi-sensor fusion is fundamental to environment perception in intelligent connected vehicles, but its reliability depends on maintaining a consistent spatiotemporal reference among cameras, LiDAR, millimetre-wave radar, and IMU/GNSS. This review examines multi-sensor calibration from a vehicle life-cycle perspective, focusing on how calibration parameters drift, how miscalibration is detected, when recalibration should be triggered, how parameters are recovered online, how the vehicle should operate while calibration is uncertain, and how recovery should be verified. Intrinsic, extrinsic, and temporal parameters are distinguished from pairwise and joint calibration architectures, while temporal errors are further separated into fixed offsets and time-varying synchronisation effects. Calibration methods are organised by their constraint sources, including artificial targets, natural-scene structure, and motion consistency, with learning-based approaches treated as an implementation paradigm rather than a separate information source. The review further compares requirements across end-of-line production, research platforms, and in-service maintenance, showing that in-service accuracy should be defined by downstream functional tolerance. A major remaining challenge is online verification: existing residuals and consistency checks provide only indirect evidence, while accepted criteria linking calibration uncertainty to downstream safety margins are still lacking. Future research should integrate detection, recovery, degraded operation, and verification within a unified maintenance loop.

Keywords: Intelligent connected vehicles, Multi-sensor calibration, Spatiotemporal calibration, Calibration drift, Online verification

1. Introduction

Intelligent connected vehicles perceive their environment through the coordinated operation of cameras, LiDAR, millimetre-wave radar and IMU/GNSS. Because no single sensor is simultaneously adequate in range measurement, semantic content and environmental robustness, multi-sensor fusion has become the mainstream route for production perception systems. Fusion holds, however, only if the measurements of all sensors are placed within a common spatiotemporal reference. Once that reference is in error, a point cloud projects onto the wrong image region and the pipeline may associate a pedestrian with empty roadway while reporting high confidence. Errors in

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.37905

the relative pose between sensors propagate directly into detection, tracking and localisation performance [1].

Critically, this error does not arise only at assembly. After a vehicle enters service, chassis vibration, mechanical wear, thermal cycling, suspension deflection and minor collisions cause calibration parameters to drift, and the drift is gradual enough to escape human notice: the system continues to operate normally while perception accuracy quietly decays, and safety is threatened if the condition is neither detected nor corrected [2]. A calibration that was accurate at delivery is therefore no guarantee of a calibration that is accurate three years later.

Sensor calibration addresses this by solving for three classes of parameters: the intrinsics describing an individual sensor's own measurement model, the extrinsics describing the spatial relationship between sensors, and the temporal parameters describing the relationship between nominal timestamps and true sampling instants. Because the downstream detection, tracking, localisation and planning modules all inherit these parameters, calibration accuracy is a foundational condition for the performance of the perception system: an error introduced at this stage is passed on to every module that follows and cannot be recovered by them. It is not the only determinant — detection and tracking models, sensor noise, and the environmental conditions under which the system operates all shape the performance finally achieved — but it sets the quality of the spatiotemporal reference on which the rest depends, which makes calibration a safety-relevant function rather than an optional preparatory step.

Research over the past two decades has advanced along several coupled directions rather than one. Dependence on dedicated calibration targets has been progressively reduced, beginning with Zhang's planar-target method and continuing through targetless approaches that exploit cues already present in natural scenes [2, 3]. In parallel, spatial and temporal parameters have moved from separate to joint estimation. Observability analysis has been introduced to characterise when a calibration problem is well constrained [4-6]. Automatic fault detection has emerged as a capability distinct from parameter estimation and cross-platform generalisation and long-term stability have become explicit evaluation criteria [1, 6].

Taken together, these directions describe a shift in what calibration is: from a measurement performed once, under controlled conditions, to a capability the vehicle must sustain over its service life. This review takes that life cycle as its organising thread and follows it as a closed loop of six questions — why parameters lose validity, how the loss is detected, when recalibration must be triggered, how the vehicle should be operated while calibration is in doubt, how parameters are recovered under the observability conditions available at the time, and how the recovered parameters are shown to be adequate for the tasks that consume them.

2. Common sensor types and their complementarity

2.1. Camera

Cameras record dense colour and texture, making them the primary source of semantic information for classification, lane marking recognition and traffic sign reading. They cannot measure range directly: depth must be inferred from stereo disparity or learned priors, both of which degrade at distance. Performance also depends strongly on illumination, and low sun angles, tunnel entrances, and night driving all reduce reliability.

DOI: 10.54254/2753-8818/2026.37905

2.2. LiDAR

LiDAR returns direct three-dimensional range measurements with centimetre-level accuracy and is largely insensitive to ambient lighting. Its point clouds are geometrically precise but semantically sparse, and angular resolution falls with distance. Return intensity is attenuated by rain, fog and airborne dust, which limits availability in adverse weather.

2.3. Millimetre-wave radar

Radar measures radial velocity directly through the Doppler effect and penetrates rain, fog and dust, giving it the best all-weather availability of the four modalities. Its point clouds are sparse and noisy, elevation resolution is poor in conventional units, and multipath returns produce spurious detections. Recent 4D radar adds elevation and denser returns, but sparse, noisy clouds and unintuitive spectral data still make radar hard to calibrate against other modalities [7-9].

2.4. IMU and GNSS

An IMU measures angular rate and specific force at hundreds of hertz, providing ego-motion at a rate far above that of the exteroceptive sensors, but its integrated pose drifts without external correction. Where GNSS is available it bounds that drift and can additionally supply a timing reference, since its pulse-per-second output is disciplined by satellite atomic clocks. Whether that signal actually serves as the system time base is a matter of architecture rather than a general property: some platforms distribute GNSS PPS to all sensors, others synchronise to a vehicle-internal clock or align timestamps in software, and GNSS is in any case degraded in tunnels and urban canyons.

2.5. Complementarity and the resulting calibration requirements

These modalities are complementary in precisely the dimensions in which each is individually deficient, and every complementary pairing generates a requirement on the spatiotemporal reference. It is worth stating that requirement carefully, because two distinct propositions are easily conflated. The first is that a production system must establish a consistent set of spatial relations among all sensors, so that any measurement can be expressed in any frame the downstream modules use. The second, stronger and not generally necessary, is that every sensor pair must be calibrated directly. Where each sensor is accurately referenced to a common vehicle or body frame, the relative transformation between any two of them follows by composition, and no dedicated procedure for that pair is required. Toolboxes addressing this sensor-to-vehicle problem have appeared alongside the sensor-to-sensor literature [10].

What decides between direct calibration and composition is not principle but error budget and observability. Composition inherits the uncertainty of every transformation on the path, so a pairing whose residual error is consumed directly by a tightly coupled fusion stage — camera–LiDAR projection being the usual example — is often calibrated directly even when it could in principle be composed, whereas pairings used only loosely, or for which no direct procedure is practical, are more readily referenced through the body frame. A useful way to describe a production suite is therefore as a sensor graph that must be connected and consistent, in which the choice of which edges to measure directly is a design decision. The edges are not equivalent in kind, either. Camera–LiDAR admits direct geometric correspondence between features seen in both modalities. An IMU

DOI: 10.54254/2753-8818/2026.37905

observes no external scene at all, so no cross-modal spatial feature correspondence can be constructed between it and an exteroceptive sensor. Such edges must instead rest on the consistency of the motion each sensor independently reports.

3. Fundamental problems in sensor calibration

3.1. Parameter types: what is estimated

3.1.1. Intrinsic parameters

Intrinsic calibration recovers the measurement model internal to one sensor. For a camera this comprises focal length, principal point and distortion coefficients, which map a three-dimensional point in the camera frame to a pixel. Zhang's method [3] remains standard, requiring only a planar pattern observed from several orientations. For LiDAR the intrinsics are per-beam angular and range offsets, estimable by minimising point-to-local-plane error over the returned cloud [11].

3.1.2. Extrinsic parameters

Extrinsic calibration recovers the six-degree-of-freedom rigid transformation between two sensor frames. Its error is range-dependent: a rotation error of 0.1° displaces a projected point by roughly 9 cm at 50 m, so deviations far below the threshold of visual inspection can move a distant object out of its correct image region. Translation error, by contrast, produces a displacement that is largely independent of range but dominates at short distances, which is one reason a single scalar extrinsic error is a poor summary of the quantity that matters downstream.

3.1.3. Temporal parameters: a fixed offset and a family of time-varying errors

Temporal calibration is frequently reduced to a single constant offset between two sensor streams. That reduction is convenient and, for an offline estimate on a short sequence, often adequate, but it does not describe the temporal error present in a vehicle over hours of operation. Two problems are involved, and they are not the same problem.

The first is the fixed offset: a constant difference between the timestamp a sensor reports and the instant at which the measurement was actually taken, arising from a deterministic trigger delay, a fixed transmission and processing latency, or a constant bias between two clocks. It is a single parameter per stream or per pair; it is estimable jointly with the extrinsics, and once estimated, it can be applied as a correction [10].

The second is a family of time-varying synchronisation errors, which no single parameter absorbs. Independent oscillators differ not only in phase but in frequency, so an uncorrected clock skew accumulates as a slowly growing offset and must be re-estimated or continuously tracked. Communication latency and buffering jitter vary with bus load and with the state of the processing pipeline, so the delay of a frame is a distribution rather than a constant. Within a single measurement, sampling is not instantaneous: a rolling-shutter camera exposes image rows sequentially, so the correct timestamp differs from row to row, and a mechanically scanning LiDAR samples the points of one revolution over the whole scan period, so points in the same frame correspond to different vehicle poses and require per-point timing to be de-skewed correctly. Continuous-time trajectory representations address exactly this regime, since any pose can be interpolated at a measurement's own timestamp without introducing a variable per measurement [8].

DOI: 10.54254/2753-8818/2026.37905

The magnitudes involved are not negligible: at 60 km/h a 50 ms error corresponds to 0.83 m of displacement, and a full LiDAR revolution at 10 Hz spans 100 ms of vehicle motion. Hardware synchronisation bounds several of these effects at source where the architecture supports it — a common trigger removes the fixed offset and constrains skew — but it does not remove intra-measurement timing, and it is unavailable on platforms that align timestamps in software. A paper that claims spatiotemporal calibration should therefore state which of these effects it estimates. Treating the temporal parameter as one number leaves the time-varying part of the problem unaddressed rather than solved.

3.2. Coupling among the three types

The three are not independent. An uncompensated temporal error biases extrinsic estimation on a moving platform, since the sensors then observe from different poses, and the bias is systematic rather than averaging out — which is why residual alone is a poor diagnostic when timing is uncertain. Intrinsic error likewise propagates into extrinsic residuals, where an optimiser will partly absorb it into the pose it returns.

3.2.1. Calibration architecture: an orthogonal axis

Whether calibration is performed pairwise or jointly is a question of system organisation, not of parameter type. The two axes are orthogonal: any architecture may estimate any parameter type.

3.2.2. Pairwise inconsistency and transform-chain error propagation

Two distinct problems follow from estimating transformations independently, and they are often conflated. The first is pairwise inconsistency: independent estimates need not satisfy the loop constraints of the sensor graph, so composing camera–LiDAR with LiDAR–IMU will not in general reproduce a separately estimated camera–IMU transformation. The loop residual is a property of the parameter set itself, does not grow with anything, and is computable without ground truth.

The second is transform-chain error propagation, arising at run time when a measurement is propagated through several transformations to reach a target frame. Every additional transformation introduces the uncertainty of one more estimated parameter set, so a longer chain draws on more estimated quantities and admits more opportunity for error to accumulate. Keeping chains short is sound practice for that reason. It does not follow, however, that a longer chain necessarily yields a larger error, still less that the error increases monotonically with the number of links. What propagates is a covariance, not a scalar: the contribution of each link depends on the direction of its error relative to the others, on the lever arms involved, and on whether the estimates are statistically independent. Errors estimated in a common optimisation are typically correlated, and correlated errors may partially cancel — a chain calibrated jointly can be more accurate at its endpoints than one whose links were estimated separately, even with more links. The defensible statement is that each composition adds parameter uncertainty and creates the *opportunity* for accumulation, with the resulting magnitude set by the covariance propagation of the specific graph rather than by the number of links. The first problem, then, is a consistency defect of the parameter set. the second is an accuracy penalty incurred at run time.

DOI: 10.54254/2753-8818/2026.37905

3.2.3. Joint calibration, the common reference frame, and observability

Joint calibration estimates all parameters in one optimisation. Enforcing loop consistency is one of its effects, but describing it only in those terms understates what the formulation involves, and overstates what it delivers.

Three issues arise that pairwise estimation can avoid or defer. The first is the choice of common reference frame. A production system requires each sensor to be referenced to the vehicle body frame, both because planning and control operate in body coordinates and because direct referencing shortens the chains discussed above. The body frame is itself a convention realised through a physical datum, and how it is defined affects what the estimated parameters mean [11]. The second is gauge freedom. A sensor graph constrained only by inter-sensor measurements determines the parameters up to a global transformation, so unless the datum is fixed — by anchoring one sensor, by adding body-frame measurements, or by a prior — the optimisation is singular in those directions, and an apparently converged solution may simply have settled wherever the initialisation placed it. The third is residual weighting. Reprojection error in pixels, point-to-plane distance in metres and inertial residuals in their own units are not commensurable, so the solution depends on the relative weights assigned, which should follow from each sensor's noise model rather than from convenience. a mis-specified weighting biases the estimate towards whichever sensor is implicitly trusted most.

Underlying all three is observability, which is where joint calibration is most often misread. Solving all parameters together does not create information that the data do not contain. If the motion is insufficiently exciting or the scene insufficiently constraining, certain parameter directions remain unobservable, and the optimiser will return values in those directions that are determined by the prior, the regularisation or the initialisation rather than by the measurements. Such a solution can be perfectly loop-consistent and still wrong: consistency constrains the parameters relative to one another, not relative to the physical configuration. This is the reason observability analysis has moved from a post-hoc check to a design consideration — identifying the motions under which LiDAR–IMU calibration degenerates, and by extension deciding when an online estimate should be accepted at all. Joint calibration and observability analysis are best understood as a pair: the first pools the available constraints, the second states which parameters those constraints actually determine.

3.2.4. Drift: shared causes, parameter-specific sensitivities

Drift is not one phenomenon acting uniformly on all three parameter types, and treating it as generic mechanical wear obscures the diagnostic problem. It is equally inaccurate, though, to assign one physical cause to each parameter type. The physical causes present in a vehicle — mechanical load, thermal cycling, ageing — act on the whole sensing assembly, and each parameter type responds to several of them. What differs is which cause each type is *most sensitive* to, and it is that ranking, rather than an exclusive mapping, which supports diagnosis.

Extrinsics respond most directly to mechanical disturbance. Vibration, a minor collision or a loosening bracket alters the geometry between sensor housings; small mechanical changes have disproportionate downstream effects. Extrinsics are not, however, insensitive to temperature: mounting brackets, roof rails and windscreen fixtures expand and contract with ambient and engine-bay temperature, so a suite can exhibit a reversible diurnal or warm-up variation in the sensor-to-sensor transformation with no mechanical change whatever. A monitor that attributes every extrinsic

DOI: 10.54254/2753-8818/2026.37905

deviation to mechanical damage will misdiagnose this pattern, and a maintenance policy that triggers a workshop visit on it will trigger repeatedly and unnecessarily.

Intrinsics respond most strongly to thermal effects: expansion of the lens barrel changes effective focal length and shifts the principal point, with comparable temperature dependence reported for LiDAR range offsets. Yet mechanical events reach the interior of a sensor as well. Shock and sustained vibration can displace lens elements or the image sensor within their mounts, and an impact severe enough to move a bracket may also disturb the internal optomechanical alignment of the device attached to it. Intrinsic and extrinsic changes can therefore occur together and from a single event, which complicates attribution: a residual increase after a kerb strike does not identify which of the two changed.

Temporal parameters are governed principally by the timing chain — oscillator behaviour, link reconfiguration, variable buffering under load — but the oscillators themselves are temperature-dependent components, and the frequency error of an uncompensated crystal varies measurably over the temperature range a vehicle experiences. Thermal variation can thus appear in all three parameter types at once, through entirely different physical paths.

Two consequences follow. First, the observable signature of a deviation, not its assumed cause, is the practical basis for diagnosis: mechanical events tend to produce step changes that persist, thermal effects a slow variation correlated with temperature and at least partly reversible on cooling, and timing-chain effects a variation correlated with computational and communication load. Second, because more than one parameter type can move at once and from the same event, a monitor tuned to one signature will not detect the others, and an estimator that corrects one type while holding the others fixed may absorb another type's error into the parameters it does adjust.

4. Calibration methods as constraint sources

4.1. Basis of the taxonomy

The methods reviewed here are the means by which each stage of the maintenance loop obtains its constraints, and the taxonomy used is organised on that basis. The primary axis is the *source of the constraint*: whether the information that determines the parameters comes from an artificial object of known geometry placed in the shared field of view, from structure already present in the natural scene, or from the consistency of the motion each sensor independently reports. These three are the primary classification framework adopted in this paper, not a claimed partition of every possible source of constraint. Hybrid procedures combine them, and other sources exist that none of the three describes cleanly — a prior map or previously built point cloud against which a sensor is registered, a CAD or kinematic model of the vehicle supplying nominal mounting geometry, and manufacturing or assembly tolerances used as priors. The three are taken as the primary axis because they organise the bulk of the literature and because they differ in exactly the respects that matter operationally: whether the vehicle must be taken out of service, whether the sensors must observe common structure, and what must be true of the scene or trajectory for the parameters to be observable.

A second and independent axis is *how* the estimator is implemented: by explicit geometric modelling and numerical optimisation, or by a function learned from data. Learning is not a fourth source of constraint. A learned network still derives its information from targets, from scene structure or from motion. What it changes is how the mapping from that information to the parameters is obtained, and what is required to obtain it. Because the two axes are independent, a learned method may consume any of the three constraint sources, and the two are treated separately

DOI: 10.54254/2753-8818/2026.37905

here for that reason: Earlier surveys have organised the field mainly by sensor pair or by algorithmic principle [1, 2]. Separating supply from implementation is intended to make clear which properties of a method follow from the information it uses and which from how it is built.

Even within the constraint-source axis, membership is not exclusive in practice. A published system may use a target to initialise, and scene features to refine, or estimate trajectories from scene features and then impose motion consistency. The columns of Table 1 should therefore be read as the sources a method draws on, not as disjoint families of systems.

4.2. Target-based methods

Target-based methods place an artificial object of known geometry in the shared field of view and establish explicit correspondences from it. For camera intrinsics, planar checkerboards remain standard [3]. For camera–LiDAR pairs, targets are designed so both modalities can localise the same features: perforated boards recovered simultaneously in image and point cloud, spherical targets whose centre is view-independent, and ordinary boxes whose orthogonal planes constrain rotation. For camera–radar and LiDAR–radar, the target must also be radar-visible; in practice, a metallic corner reflector with high radar cross-section is used. Static multi-target arrangements have calibrated camera, radar and LiDAR simultaneously against a common reflector set.

The reason this route is associated with high accuracy is structural: the geometry of the target is known independently of the estimate, the correspondences are unambiguous, and the observation geometry can be chosen to make the parameters well conditioned rather than accepted as the scene provides it. The accuracy actually achieved still depends on how precisely each modality localises the target — a corner reflector is localised far less sharply by a sparse radar than a checkerboard corner is by a camera — and on target placement, so the margin over targetless routes varies with the sensor pair. The characteristic cost is operational rather than computational: a controlled site, a trained operator and vehicle downtime.

4.3. Scene-feature-based methods

Targetless methods exploit structure already present in the environment. Information-theoretic approaches maximise the mutual information between LiDAR reflectivity and image intensity, exploiting the statistical dependence between the two without explicit correspondences. Geometric approaches extract edges, lines or planes visible in both modalities and minimise a reprojection residual. Edge-based formulations report pixel-level alignment in unmodified environments, and line-based variants exploit urban straight structure. Semantic approaches align segmentation outputs, but aligning discrete masks yields a strongly non-convex objective sensitive to segmentation noise and occlusion [1].

These methods drop the target but retain the need for direct cross-modal correspondence, and therefore for overlapping observation of the same structure. Their accuracy is conditional on the scene supplying it: performance degrades on open highway or in any scene lacking co-observed geometry, and moving objects introduce correspondences that are valid at one instant but not across an estimation window. Because the objective is non-convex, the quality of the initial estimate matters, which is one reason this route is well suited to refinement from an approximately correct prior — the situation of online recovery — and less suited to calibration from scratch.

DOI: 10.54254/2753-8818/2026.37905

4.4. Motion-consistency-based methods

These methods estimate each sensor's trajectory independently and solve for the transformation that renders the trajectories consistent, the classical hand-eye problem AX = XB. Their defining property is that they require no cross-modal feature correspondence, only comparable motion, which is why they are the route most commonly reported for LiDAR–IMU and camera–IMU, where an inertial sensor observes nothing of the external scene against which features could be matched. Kalibr established unified temporal and spatial estimation for camera–IMU suites, and motion-based formulations recovering extrinsics and timing offset together followed for heterogeneous pairs [10]. For LiDAR–IMU, LI-Calib represents the trajectory as a continuous-time B-spline, accommodating high-rate asynchronous measurements [8]. Later work analysed observability and identified the motions under which the problem degenerates. Camera–IMU calibration must additionally recover gravity direction and, in monocular configurations, a scale ambiguity observable only under acceleration. The same representation supports joint spatiotemporal calibration of full LiDAR/camera/IMU suites, since any pose can be interpolated on the spline at its own timestamp without adding variables.

The characteristic cost is dependence on exciting motion, and it is a hard dependence rather than a matter of degraded accuracy: planar trajectories and constant-velocity driving leave specific degrees of freedom unobservable, and in those directions the estimate is set by the prior rather than by the data, with a residual that gives no indication of the fact. This is why the route is usually paired with an observability check, and why, in service, it is more naturally applied opportunistically — when the vehicle happens to execute a trajectory that excites the parameters — than on demand.

4.5. Comparison of the three constraint sources

Table 1 compares the three constraint sources across the dimensions that determine where each can be used. Because the entries derive from different datasets, sensor configurations and evaluation protocols, and because a given system may draw on more than one source, the table summarises reported tendencies together with the conditions under which they hold. it is not a controlled benchmark, and the figures in one column are not directly comparable with those in another.

No source dominates. Target-based methods are the ones that can be made well conditioned by construction, and where a controlled site is available they are the natural reference, but they cannot run on a vehicle in service. Scene-feature methods remove the target at the price of a dependence on what the scene contains, and are strongest as refinement from a good prior. Motion-consistency methods are the only route available where no cross-modal correspondence exists at all, and the route that most often estimates temporal parameters jointly, but they are bounded by what the trajectory excites. These are properties of the information each source supplies, and they persist regardless of how the estimator is implemented.

Table 1. Reported characteristics of the three constraint sources

Dimension	Target-based	Scene-feature-based	Motion-consistency-based
Constraint source	Artificial object of known geometry	Structure present in the natural scene	Consistency of independently estimated motion
External target required	Yes	No	No
Direct cross-modal correspondence required	Yes	Yes	No

DOI: 10.54254/2753-8818/2026.37905

Table 1. (continued)

Parameters typically recovered	Intrinsics and extrinsics	Extrinsics	Extrinsics, frequently with temporal parameters
Conditions under which reported accuracy holds	Sub-centimetre and sub-degree figures reported on controlled setups. Depends on how sharply each modality localises the target, so the margin over targetless routes varies by sensor pair	Pixel-level alignment reported for camera–LiDAR in structured scenes. degrades where the scene supplies little co-observed structure	Varies strongly with trajectory excitation. under degenerate motion, the affected directions are unobservable and the residual does not bound the error
Operating stage as usually deployed	Offline	Offline, or online as refinement from a prior	Offline batch, or online sliding window
Sensitivity to dynamic scenes	Low (controlled site)	High. Moving objects generate correspondences valid only instantaneously	Moderate. moving objects bias the estimated trajectories
Reported computational profile	Low per solve. Cost dominated by site setup and operator labour	Generally the heaviest per solve, the objective being non-convex and evaluated over dense data, though reported cost varies widely with formulation, data volume and hardware	High for batch estimation. Sliding-window and incremental formulations reported at online rates
Sensor pairings reported	camera–LiDAR, camera–radar, LiDAR–radar	Predominantly camera–LiDAR. Also LiDAR–LiDAR and radar–camera	LiDAR–IMU, camera–IMU, camera–LiDAR, full suites

5. Learning as an implementation paradigm

Learning-based methods regress parameters from raw measurements through a function fitted to training data rather than through an explicit geometric objective solved at run time. Treated as a separate implementation axis, they are best characterised by what the substitution changes and what it leaves untouched.

RegNet first demonstrated end-to-end regression, combining feature extraction and matching in one network with iterative refinement. CalibNet added geometrically supervised losses. LCCNet replaced explicit matching with a cost-volume correlation layer [7]. Quality-aware formulations weighted unreliable correspondences explicitly, and transformer architectures improved long-range association. A related line applies self-supervision to the temporal problem, learning the synchronisation offset without timing labels. Radar–camera application is more recent and less mature, reflecting sparse returns and the absence of a dense correspondence signal [9].

Two properties recur, and both follow from the paradigm rather than from any particular architecture. The first is the cost profile: the expensive stage is training, performed once and offline, while inference is a forward pass whose cost is fixed and predictable — an advantage for an online loop that must budget compute deterministically, though the comparison with classical optimisation depends on the hardware available and on whether the network output is accepted directly or used to initialise a refinement. The second is generalisation. Accuracy degrades when the sensor model, mounting geometry or data distribution departs from the training configuration, and since a production programme varies across variants and model years, transfer is constrained more severely than single-dataset benchmark figures suggest. Reported figures should be read with their

DOI: 10.54254/2753-8818/2026.37905

configuration attached: the 1.01 cm / 0.12° reported by LCCNet on KITTI [2, 7], for instance, characterises one dataset and one sensor configuration, not a property of learned calibration as such.

What learning does not change is the information available. A network trained on scene appearance still requires the modalities to observe common structure. A learned estimator applied to motion data is still bounded by what the trajectory excites, since observability is a property of the problem and not of the estimator. Table 2 makes this mapping explicit.

Table 2. Learned implementations mapped onto the constraint sources

Constraint source consumed	Reported learned implementations	What learning changes	What it does not change
Artificial target	Little reported as end-to-end calibration. where learning appears, it is generally confined to detecting or localising the target inside an otherwise classical pipeline	Robustness of target detection under poor illumination or partial visibility	The need for a controlled site, a physical target and vehicle downtime
Natural scene structure	The main body of learned calibration: RegNet, CalibNet, LCCNet, DXQ-Net, CalibFormer. Radar–camera work emerging	Replaces hand-designed features and explicit correspondence search with a learned representation and regression. Shifts cost from run time to training	The requirement that the modalities observe common structure, and the dependence of accuracy on what the scene contains
Motion consistency	Self-supervised estimation of the synchronisation offset without timing labels. Learned odometry front-ends supplying the trajectories to be aligned	Removes the need for labelled timing or pose ground truth. tolerates front-end noise learned from data	Dependence on exciting motion. degenerate trajectories leave the same directions unobservable

5.1. The online maintenance loop

The loop introduced in Section 1 comprises six problems that are frequently treated as one. They differ in what is being asked, in what counts as a method, and in how success is measured, and progress on them is markedly uneven. The first four concern the parameters themselves. The last two concern what the vehicle is entitled to do with them, and are the two on which the literature is thinnest.

5.2. How drift arises

This paper established that the three parameter types respond to overlapping physical causes with different sensitivities, and therefore on different time scales and with different signatures. Extrinsic drift is the dominant concern in service because mechanical disturbance acts continuously. The practical difficulty is that the magnitudes of concern are small: deviations well below what an operator would notice on a projected overlay are sufficient to displace a distant object, so the regime that matters is precisely the regime that is invisible to inspection [5]. A second difficulty is that not every deviation is a fault. Reversible thermal variation and a loosening bracket can produce similar instantaneous residuals while requiring entirely different responses, and distinguishing them requires the temporal signature rather than the magnitude alone.

DOI: 10.54254/2753-8818/2026.37905

5.3. How miscalibration is detected

Detection asks a different question from estimation. Estimation asks what the parameters are. detection asks only whether the current ones remain valid, and it must answer continuously, cheaply and without a target. Recognising this asymmetry has produced a reframing from regression to classification, in which a lightweight network judges calibration state rather than recovering parameters, at a fraction of the cost of full re-estimation [6]. Other work monitors motion cues, comparing the displacement each modality reports for the same physical structure [5]. Useful metrics are the minimum detectable misalignment, the detection latency in frames or seconds, and the false-positive rate under nominal operation.

The central difficulty is discrimination. Occlusion, adverse weather and heavy dynamic traffic all degrade cross-modal agreement without any parameter having changed, and a detector that cannot separate these from genuine drift will either raise constant false alarms or be desensitised past usefulness. The consequence is that a detector output should be treated as a hypothesis with a quantified false-positive rate rather than as a fault declaration, where the vehicle must act on that hypothesis before it can be confirmed.

5.4. When recalibration is triggered

Triggering is a decision problem layered on detection. The natural threshold is not any detectable deviation but the deviation at which downstream performance degrades beyond tolerance. Two further considerations shape the policy. Hysteresis is needed so that a statistic fluctuating about the threshold does not cause repeated recalibration. And because motion-consistency methods require excitation and scene-feature methods require co-observed structure, a practical policy is opportunistic: defer the attempt until the vehicle encounters conditions under which the affected parameters are observable, rather than attempting recovery immediately on detection. The cost of that policy is that the interval between detection and correction may be long and is not under the system's control, which is precisely why the interval needs a defined operating regime of its own.

5.5. How parameters are recovered online

Recovery in service is constrained as offline calibration is: no target, no downtime, no privileged manoeuvre, only the computation left over from perception. What makes it tractable is that recovery is normally local refinement from an approximately correct prior rather than calibration from scratch, which narrows the search and improves conditioning. Targetless scene-feature refinement seeded from the previous parameters is the most direct route and frameworks combining detection with correction have been demonstrated on vehicle platforms. Continuous-time formulations support re-estimation over a sliding window, recovering spatial and temporal parameters jointly. Learned regressors are attractive for their fixed inference cost.

The condition that governs all of these is observability. An online solver will return a parameter set whether or not the data determine it, and in the unobservable directions the value returned reflects the prior rather than the measurements. An online recovery step should therefore report not only a parameter estimate but an indication of which directions the available data actually constrained — information the verification stage cannot reconstruct afterwards from the residual alone.

DOI: 10.54254/2753-8818/2026.37905

6. Conclusion

This paper reviewed multi-sensor calibration for intelligent connected vehicles, organised around the life cycle of a calibration rather than around a classification of methods. This paper identified the spatial relations a production suite must establish and distinguished the requirement for a consistent sensor graph from the assumption that every pair must be calibrated directly. This paper separated the three parameter types from the architectural question of pairwise versus joint estimation, decomposed the temporal problem into a fixed offset and a family of time-varying synchronisation errors, distinguished pairwise inconsistency from transform-chain error propagation, examined what joint calibration does and does not deliver in the absence of observability, and described drift in terms of the cause each parameter type is most sensitive to.This paper reviewed the three constraint sources on which the maintenance loop draws and treated learning separately as an implementation paradigm that may consume any of them. This paper developed the loop as six problems — how drift arises, how it is detected, when recalibration is triggered, how parameters are recovered, how the vehicle is operated while calibration is in doubt, and how recovery is verified. This paper explained why the requirements at three life-cycle stages differ in kind rather than only in magnitude.

Development is proceeding along several directions at once rather than converging on a single one. Dependence on artificial targets has been reduced but not eliminated, and remains appropriate where a controlled site exists. Spatial and temporal parameters are increasingly estimated jointly rather than sequentially, and continuous-time formulations have made the intra-measurement timing of scanning and rolling-shutter sensors tractable within the same estimate. Observability analysis has moved from an afterthought to a design consideration, determining when online estimation is admissible at all. Automatic fault detection has separated from parameter estimation and become a capability in its own right. Cross-platform generalisation and long-term stability have entered evaluation criteria alongside single-configuration accuracy. Underlying all of these is the change of role noted at the outset: calibration is becoming a function the vehicle maintains, and the questions that now limit progress are increasingly about what the vehicle may do with its parameters rather than about how to estimate them.

Four limitations stand out. First, verification of an online correction remains without an accepted criterion, the least mature stage of the loop as identified: the difficulty is not a shortage of internal residuals but the absence of an acceptance criterion expressed in terms of downstream safety tolerance, and a loop that cannot certify its own output cannot be trusted to run unsupervised. Second, the temporal problem is still largely treated as a single fixed offset, while the errors that matter in service — clock skew, load-dependent latency and jitter, and the intra-measurement sampling times of rolling-shutter and scanning sensors — vary over time and are addressed by comparatively few formulations. Third, radar-inclusive calibration lags the camera–LiDAR case substantially, and the sparsity that makes radar valuable in poor weather is the same property that makes it hard to calibrate. Fourth, the absence of shared benchmarks for online calibration means reported figures are not comparable across studies, as Table 1 reflects.

Useful next steps follow from these. Benchmarks with controlled injected drift and standardised detection metrics would make detection and verification methods comparable, and would allow an acceptance criterion to be characterised against deviations of known magnitude. Evaluation should be task-level rather than parameter-level, since the tolerable error is defined by downstream performance. Detection, recovery, degraded operation and verification, still largely studied in

DOI: 10.54254/2753-8818/2026.37905

isolation, need to be specified and evaluated as one loop, including the behaviour of the vehicle during the interval in which its calibration is in doubt.

References

[1] An, P., Ding, J., Quan, S., Yang, J., Yang, Y., Liu, Q., & Ma, J. (2024). Survey of extrinsic calibration on LiDAR-camera system for intelligent vehicle: Challenges, approaches, and trends. *IEEE Transactions on Intelligent Transportation Systems, 25*(11), 15342–15366. https://doi.org/10.1109/TITS.2024.3419758

[2] Li, X., Xiao, Y., Wang, B., Ren, H., Zhang, Y., & Ji, J. (2023). Automatic targetless LiDAR–camera calibration: A survey. *Artificial Intelligence Review, 56*(9), 9949–9987. https://doi.org/10.1007/s10462-022-10317-y

[3] Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(11), 1330–1334. https://doi.org/10.1109/34.888718

[4] Yan, G., Liu, Z., Wang, C., Shi, C., Wei, P., Cai, X., Ma, T., Liu, Z., Zhong, Z., Liu, Y., Zhao, M., Ma, Z., & Li, Y. (2022). OpenCalib: A multi-sensor calibration toolbox for autonomous driving. *Software Impacts, 14*, 100393. https://doi.org/10.1016/j.simpa.2022.100393

[5] Peng, P., Pi, D., Yin, G., Wang, Y., Xu, L., & Feng, J. (2024). Automatic miscalibration detection and correction of LiDAR and camera using motion cues. *Chinese Journal of Mechanical Engineering, 37*, 50. https://doi.org/10.1186/s10033-024-01035-3

[6] Huang, Z., Li, J., Zhao, H., Ma, X., Zhong, P., Zhou, X.-H., Ye, W., & Fan, R. (2025). Environment-driven online LiDAR-camera extrinsic calibration. *IEEE Transactions on Automation Science and Engineering, 22*, 24164–24176. https://doi.org/10.1109/TASE.2025.3627253

[7] Beltrán, J., Guindel, C., de la Escalera, A., & García, F. (2022). Automatic extrinsic calibration method for LiDAR and camera sensor setups. *IEEE Transactions on Intelligent Transportation Systems, 23*(10), 17677–17689. https://doi.org/10.1109/TITS.2022.3155228

[8] Lv, J., Zuo, X., Hu, K., Xu, J., Huang, G., & Liu, Y. (2022). Observability-aware intrinsic and extrinsic calibration of LiDAR-IMU systems. *IEEE Transactions on Robotics, 38*(6), 3734–3753. https://doi.org/10.1109/TRO.2022.3174476

[9] Li, S., Li, X., Chen, S., Zhou, Y., & Wang, S. (2024). Two-step LiDAR/camera/IMU spatial and temporal calibration based on continuous-time trajectory estimation. *IEEE Transactions on Industrial Electronics, 71*(3), 3182–3191. https://doi.org/10.1109/TIE.2023.3270506

[10] Agrawal, S., Bhanderi, S., Doycheva, K., & Elger, G. (2023). Static multitarget-based autocalibration of RGB cameras, 3-D radar, and 3-D lidar sensors. *IEEE Sensors Journal, 23*(18), 21493–21505. https://doi.org/10.1109/JSEN.2023.3300957

[11] Yuan, K., Ding, L., Abdelfattah, M., & Wang, Z. J. (2022). LiCaS3: A simple LiDAR–camera self-supervised synchronization method. *IEEE Transactions on Robotics, 38*(5), 3203–3218. https://doi.org/10.1109/TRO.2022.3167455