INSTITUTE FOR PHYSICAL AI @ BMI
The Charlot Lab

Technical Report TR–2026–08
Survey / Review · Preprint v1
7 July 2026

OMNISENSE

OmniSense: The World Model as Projected Perception in a Volume

A projective account of the world model, in which where-and-what a system is emerges from a mesh of optical and radio-frequency nodes resolving a volume into occupied, empty, and unknown.

David Jean Charlot, PhD

Dean of Physical AI · The Charlot Lab, Institute for Physical AI @ BMI

Correspondence: contact@physicalai-bmi.org · physicalai-bmi.org
Interactive companion: physicalai-bmi.org/research/charlot-lab#topic · Code: github.com/dcharlot-physicalai-bmi/omnisense

Abstract. This report advances a projective account of a Physical AI system's world model. The model, including the system's own pose and identity within it, is treated not as an internal state read out from onboard sensors alone, but as a field defined over a volume by a mesh of sensing nodes: some carried by the system, some on peer systems, some anchored in the space. The mesh resolves the volume into three states. Positive space is occupied surface, reconstructed by line-of-sight optical sensing (monocular and stereo vision, LiDAR, and event cameras). Negative space is confirmed empty, carved along each clear ray from sensor to hit. The remainder is unknown, and it is concentrated behind barriers that block light. Spatial radio-frequency sensing (Wi-Fi channel state, ultra-wideband, and millimeter-wave radar) penetrates many of those barriers and resolves the space behind them, occupied and empty alike. Section 2 states the review method. Section 3 develops the three-state decomposition with occupancy log-odds and ray carving. Sections 4 and 5 survey the optical and RF pathways. Section 6 treats the distributed mesh, in which a system perceives itself from the outside through peers and anchors, connecting to cooperative-perception work. Section 7 states the position: the design objective is to drive the unknown volume toward zero, and to do so with a heterogeneous sensor set rather than any single modality. The report reports no new experimental measurements.

1. Introduction

An embodied system does not begin with a world and add sensors to it. It begins with sensors and constructs a world. The distinction matters because it changes what a world model is taken to be. On the common reading, the model is an internal representation the system maintains and updates; the body sits inside it, and perception fills it in. This report takes a different starting point. The model is a field defined over a region of space, and it is produced by projecting sensor evidence outward into that region. What the system knows about a point in the room is whatever the available sensing has established about that point, and nothing more. Where sensing has reached, the field is defined; where it has not, the field is unknown.

On this reading the system's own state is not privileged. Its pose, and in a multi-agent setting its identity relative to its peers, is one more quantity resolved by the same projection. A system carrying only body-fixed sensors perceives itself indirectly, by inference from how the world moves past it. A system embedded in a mesh of external nodes can be perceived directly, from the outside, the way a motion-capture stage perceives an actor. Both are the same operation at different node counts. The claim of this report is that the useful unit of analysis is neither the sensor nor the agent but the volume, and the useful question is how much of that volume the mesh has resolved.

The engineering content of that question is old and well developed under other names. Occupancy grids and their probabilistic update rule date to the middle 1980s[1,2]. Volumetric surface reconstruction from range data, three-dimensional occupancy mapping, and dense real-time fusion are mature[3,4,5]. Sensing through walls with radio has moved from a curiosity to a measured capability[11,12]. What this report contributes is not a new algorithm but a frame that places these under one account and names the quantity they share: the unknown fraction of a volume, and the modalities that reduce it.

2. Scope and method

This is a review and a research position. It surveys published methods for building spatial models from optical and radio sensing, and it argues for a particular way of organizing them. It reports no original experiments, no benchmark numbers of its own, and no proprietary results. Where a numerical claim appears it is attributed to a cited source. Coverage is deliberately broad across modalities rather than deep in any one, and the reader is directed to the cited work for depth. The report is explicit where a capability is early-stage: radio-frequency reconstruction of human pose and of occupancy behind barriers is a laboratory and dataset-scale result under controlled conditions, not a deployed, all-conditions guarantee, and the projective account of self-localization from a mesh is stated as a design position rather than a demonstrated system. The `omnisense` repository referenced above is a working environment for the decomposition described here, not evidence for any performance claim. The mathematics in Section 3 is standard and is included to make the three-state decomposition precise, not to report a result.

3. Three states of a volume

Partition the region of interest into volume elements. At time t each element n carries a belief that it is occupied, written as an occupancy probability $p(n)$. It is convenient to hold this belief in log-odds form,

$$L(n) = \log \frac{p(n)}{1 - p(n)},$$

because independent observations then combine by addition. Given a new measurement z_t, the recursive update is

$$L(n \mid z_{1:t}) = L(n \mid z_{1:t-1}) + L(n \mid z_t) - L_0,$$

with L_0 the prior log-odds; this is the standard occupancy filter behind grid mapping and its octree-based three-dimensional form[2,4]. The measurement term is where the geometry enters. A range sensor that returns a hit at distance r along a ray does two things at once. It raises the occupancy log-odds of the element at the hit, and, along the whole segment of the ray from the sensor to just short of the hit, it lowers the occupancy log-odds of every element the ray passed through, because a clear return is evidence those elements were empty. This second action is ray carving, and it is the mechanism that produces confirmed empty space rather than merely unmodeled space.

Three states follow directly. **Positive space** is the set of elements whose log-odds has been driven high by hits: occupied surface. **Negative space** is the set driven low by carving: confirmed empty, and therefore traversable. The **unknown** is everything left near the prior, neither hit nor carved. The value of separating negative space from unknown is operational. A planner may pass through confirmed-empty space; it may not assume the unknown is empty. For surface geometry specifically, the signed-distance formulation makes the same partition continuous: a truncated signed distance function stores, at each element, the distance to the nearest surface with sign indicating inside or outside, and the zero crossing is the reconstructed surface[3,5]. The truncation band is exactly the region the sensor has resolved; beyond it the function is undefined, which is the unknown by another name.

The unknown is not uniformly distributed. It collects behind occluders, because an optical ray stops at the first opaque surface it meets and carries no evidence about what lies beyond. Figure 1 shows the decomposition and makes this concentration visible: the shadow volume behind a barrier is where an optical-only mesh is blind. The remainder of the report is organized around that fact. Section 4 covers the modalities that build positive and negative space by line of sight. Section 5 covers the modalities that reach into the shadow.

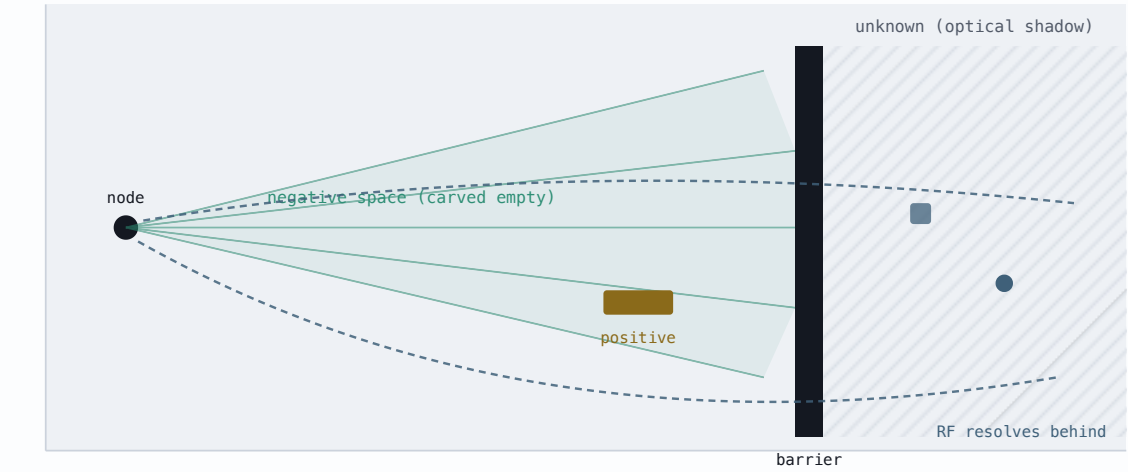

Figure 1. One volume, three states. Line-of-sight rays (green) from a node carve confirmed-empty *negative space* and terminate on occupied *positive space* (gold, and the barrier). Behind the barrier lies the *unknown*, the optical shadow (hatched). Radio-frequency paths (blue, dashed) pass through the barrier and resolve occupancy and emptiness within the shadow. An interactive version is at physicalai-bmi.org/research/charlot-lab.

Table 1. Modalities by the state they resolve and the barrier they respect or penetrate.

MODALITY	PRIMARY EVIDENCE	RESOLVES	SEES THROUGH OPAQUE BARRIERS
Monocular / stereo vision	Photometric, triangulated depth	positive + negative (line of sight)	No
LiDAR	Time-of-flight range per beam	positive + negative (line of sight)	No
Event camera	Per-pixel brightness change	positive, at motion, microsecond latency	No
Wi-Fi channel state	Multipath amplitude / phase	presence, motion, coarse pose behind walls	Yes (many materials)
Ultra-wideband	Time-of-flight ranging	range / position between nodes	Partial
Millimeter-wave radar	Range-Doppler, angle	occupancy, motion, coarse imaging	Partial

4. Line-of-sight reconstruction

The optical modalities share a geometry and differ in how they obtain range. Passive vision recovers depth by triangulation: a stereo pair, or a single moving camera over time, matches features across views and inverts the projection. The output feeds the same volumetric fusion described in Section 3, and the mature demonstrations of dense, real-time surface

reconstruction from a moving depth camera established that a running signed-distance volume can be maintained on commodity hardware[5], building on the volumetric range-image method that preceded it[3]. LiDAR obtains range directly by time of flight per beam, which makes carving unusually clean: each return is a hit at a measured distance and a carved segment behind it, and real-time LiDAR odometry and mapping is a settled technique[6]. For three-dimensional occupancy specifically, octree-backed mapping stores the log-odds field at multiple resolutions and prunes confirmed-homogeneous regions, which is what makes carving over a large volume tractable[4].

Event cameras complete the optical set from a different direction. Rather than integrating a frame, each pixel independently reports the times at which its brightness changes, at microsecond latency and high dynamic range[7]. This suits the positive-space problem under motion and difficult lighting, where a conventional frame is blurred or saturated: the surfaces that matter are exactly the ones whose projection is changing. The common limitation of the whole family is the one that motivates the next section. Every optical ray stops at the first opaque surface. Positive and negative space accumulate only within the line-of-sight cone, and the unknown behind any occluder is, to light, permanent.

5. Seeing through barriers with spatial RF

Radio-frequency signals at the wavelengths used for communication and short-range radar pass through many materials that stop visible light: drywall, wood, most furniture. This is the physical basis for reaching into the optical shadow. Three sub-modalities are relevant. Wi-Fi channel state information exposes the amplitude and phase of the multipath channel between transmitter and receiver; because a human body perturbs that channel, the channel state carries information about presence, motion, and, with learned models, coarse pose, and this has grown into a recognized sensing field[10]. The clearest demonstration that the shadow is addressable is the reconstruction of dense human pose from Wi-Fi signals alone[12] and, using dedicated radio hardware, the estimation of human pose through a wall[11]. Ultra-wideband provides precise time-of-flight ranging between nodes, which is the backbone of anchored positioning and of the mesh geometry discussed in Section 6[9]. Millimeter-wave radar returns range, Doppler, and angle, and supports occupancy detection, motion sensing, and coarse imaging, with portable three-dimensional imaging demonstrated[8].

The register of these results should be stated plainly. RF sensing does not reconstruct surfaces at optical resolution; it resolves presence, gross geometry, and motion at coarse spatial scale, and the strongest human-sensing results are obtained with trained models on curated data under controlled conditions rather than as universal, material-agnostic guarantees. That is enough for the purpose here. The three-state decomposition does not require RF to match a camera. It requires RF to convert unknown elements behind a barrier into occupied or empty ones with usable confidence, which is precisely a log-odds update from a second modality applied where the first could not reach. Positive and negative space thereby extend past the surfaces that bound the optical cone.

> The two families are complementary in a specific way. Optical sensing gives high-resolution positive and negative space wherever there is line of sight and stops at the first opaque surface. Spatial RF gives coarse positive and negative space wherever radio propagates and is largely indifferent to that surface. Neither pays for the other's dominant blindness, and the union of the two covers more of the volume than either alone.

6. The distributed mesh: perceiving yourself from the outside

Nothing in Sections 3 through 5 requires the sensors to be on one body. The log-odds field is defined over the volume, not over a device, so measurements from different vantage points combine by the same addition, provided they share a frame. This is the mesh. Its nodes are of three kinds: sensors carried by the system itself, sensors on peer systems moving through the same space, and sensors anchored in the environment. Each node contributes rays, and each ray carves and hits in the common volume. The unknown fraction is then a property of the whole mesh, and it shrinks as vantage points are added, because an occluder that shadows one node is often in full view of another.

The consequence for self-perception is the point of the report. A system with only body-fixed sensors localizes itself indirectly, inferring its own motion from the apparent motion of the world, which fails exactly when the world nearby is featureless or occluded. A system observed by peer and anchored nodes is localized directly: it appears in their positive space as an occupied, moving object, and ultra-wideband ranging between nodes fixes the metric geometry that ties their frames together[9]. The system perceives itself not from within but from the mesh at once. This is the same construction that the cooperative-perception literature builds for road vehicles, where agents share observations over a communication link to see past their individual occlusions; benchmark datasets and fusion pipelines for vehicle-to-vehicle perception show measurable gains from exactly this pooling of vantage points[13,14]. The generalization here is only that the shared quantity is the three-state field, and that one of the objects it resolves is the observing system itself.

7. Position: driving the unknown to zero

The account yields a single design objective. If the world model is the resolved fraction of a volume, then the aim of a perception system is to minimize the unknown fraction, subject to the power, bandwidth, and privacy budgets of the deployment. This reframes several familiar choices. Adding a sensor is worthwhile in proportion to the unknown volume it converts, not its raw resolution; a coarse RF node that resolves a large optical shadow can reduce the unknown more than a second camera that only densifies an already-seen surface. Active perception, the deliberate movement of a node to a new vantage, is the search for the placement that carves the most remaining unknown. Cooperation is worthwhile when a peer's cone covers one's own shadow. Each of these is a term in the same minimization.

Two boundaries keep the objective honest. First, the unknown cannot be driven to zero everywhere; sealed volumes, radio-opaque materials, and range limits leave a residue, and the

correct behavior toward that residue is to mark it unknown rather than to assume it empty, which is why the separation of negative space from unknown in Section 3 is load-bearing. Second, the modalities that see through walls also see through them into spaces where seeing may not be wanted, which places the RF pathway squarely within the governance concerns treated in the Institute's work on perception, privacy, and policy; the capability and its constraints are not separable. The projective account also connects downward to the contact layer: where no ray and no radio path reaches, the remaining route to resolving an element is to touch it, and tactile sensing closes the last of the unknown at the point of contact. OmniSense names the whole of this, the union of projected optical, radio, and contact evidence over one volume, driving the unknown toward zero.

8. Conclusion

This report described a world model as projected perception: a three-state field over a volume, produced by a mesh of nodes, in which occupied positive space and confirmed-empty negative space are carved and hit along rays, and the unknown behind barriers is resolved by radio that light cannot follow. The account rests on mature and verifiable components, occupancy log-odds and ray carving, volumetric and octree mapping, LiDAR and event sensing, Wi-Fi, ultra-wideband, and millimeter-wave sensing, and distributed cooperative perception, and its contribution is the frame that unifies them and the objective it implies. A system perceives itself from the whole mesh at once, and the work of perception is to drive the unknown fraction of its volume toward zero. The empirical validation of the projective account as a running system, and the quantification of how each modality reduces the unknown under realistic conditions, remain open and are the subject of the companion.

References

1. H. Moravec, A. Elfes. High resolution maps from wide angle sonar. Proc. IEEE Int. Conf. Robotics and Automation (ICRA), 1985. doi:10.1109/ROBOT.1985.1087316.
2. A. Elfes. Using occupancy grids for mobile robot perception and navigation. Computer 22(6), 1989. doi:10.1109/2.30720.
3. B. Curless, M. Levoy. A volumetric method for building complex models from range images. SIGGRAPH, 1996. doi:10.1145/237170.237269.
4. A. Hornung, K. M. Wurm, M. Bennewitz, C. Stachniss, W. Burgard. OctoMap: an efficient probabilistic 3D mapping framework based on octrees. Autonomous Robots 34(3), 2013. doi:10.1007/s10514-012-9321-0.
5. R. A. Newcombe, S. Izadi, O. Hilliges, et al. KinectFusion: real-time dense surface mapping and tracking. IEEE Int. Symp. Mixed and Augmented Reality (ISMAR), 2011. doi:10.1109/ISMAR.2011.6092378.
6. J. Zhang, S. Singh. LOAM: Lidar odometry and mapping in real-time. Robotics: Science and Systems (RSS), 2014. doi:10.15607/RSS.2014.X.007.
7. G. Gallego, T. Delbrück, G. Orchard, et al. Event-based vision: a survey. IEEE Trans. Pattern Analysis and Machine Intelligence 44(1), 2022. doi:10.1109/TPAMI.2020.3008413.
8. J. Moll, P. Schöps, V. Krozer, et al. Portable millimeter wave 3D imaging radar. IEEE Radar Conference, 2017. doi:10.1109/RADAR.2017.7944216.
9. S. Gezici, Z. Tian, G. B. Giannakis, et al. Localization via ultra-wideband radios: a look at positioning aspects for future sensor networks. IEEE Signal Processing Magazine 22(4), 2005. doi:10.1109/MSP.2005.1458289.

10. Y. Ma, G. Zhou, S. Wang. WiFi sensing with channel state information: a survey. ACM Computing Surveys 52(3), 2019. doi:10.1145/3310194.
11. M. Zhao, T. Li, M. Abu Alsheikh, et al. Through-wall human pose estimation using radio signals. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), 2018. doi:10.1109/CVPR.2018.00768.
12. J. Geng, D. Huang, F. De la Torre. DensePose from WiFi. arXiv:2301.00250, 2023.
13. T.-H. Wang, S. Manivasagam, M. Liang, et al. V2VNet: vehicle-to-vehicle communication for joint perception and prediction. European Conf. Computer Vision (ECCV), 2020. doi:10.1007/978-3-030-58536-5_36.
14. R. Xu, H. Xiang, X. Xia, et al. OPV2V: an open benchmark dataset and fusion pipeline for perception with vehicle-to-vehicle communication. IEEE Int. Conf. Robotics and Automation (ICRA), 2022. doi:10.1109/ICRA46639.2022.9812038.

AI-use disclosure. Preparation of this report used a large language model (Claude, Anthropic) for drafting and editing text, organizing the reviewed literature, and preparing the figures and the interactive companion. Cited references were checked to resolve to their sources. The author reviewed the content and is solely responsible for it. Consistent with ICMJE, COPE, and IEEE guidance, the model is a tool and is not credited as an author.

Institute for Physical AI @ BMI · The Charlot Lab
503 McKeever Rd, Arcola, TX 77583, USA
physicalai-bmi.org · contact@physicalai-bmi.org

Technical Report TR-2026-08 · Preprint v1