

Acquisition objective, evaluated quality and target context

FisherRF estimates parameter information from rendering derivatives; its version 2, Sections 3.1–3.4, also treats batches and mapping paths. Its revised title is “Active View Selection and Mapping with Radiance Fields using Fisher Information”; the bibliography retains the original paper identity. Rendering and completeness metrics evaluate the outcome, rather than define the acquisition objective.

Next Best Sense extends Fisher acquisition to depth and RGB during robotic Gaussian reconstruction. Acquired observations update the map online. Its object-directed tabletop setting retains modeled surroundings, while object masking is explicit for touch; the inspected method does not establish that every visual acquisition uses the same target gate. Initialization and depth supervision also affect the reported quality, separately from selected views [1]. See version 2, Sections II-B, III-B–C and IV.

The version-four reconstruction component in Figure 1 illustrates why reconstruction improvements must be separated from acquisition-score effects.

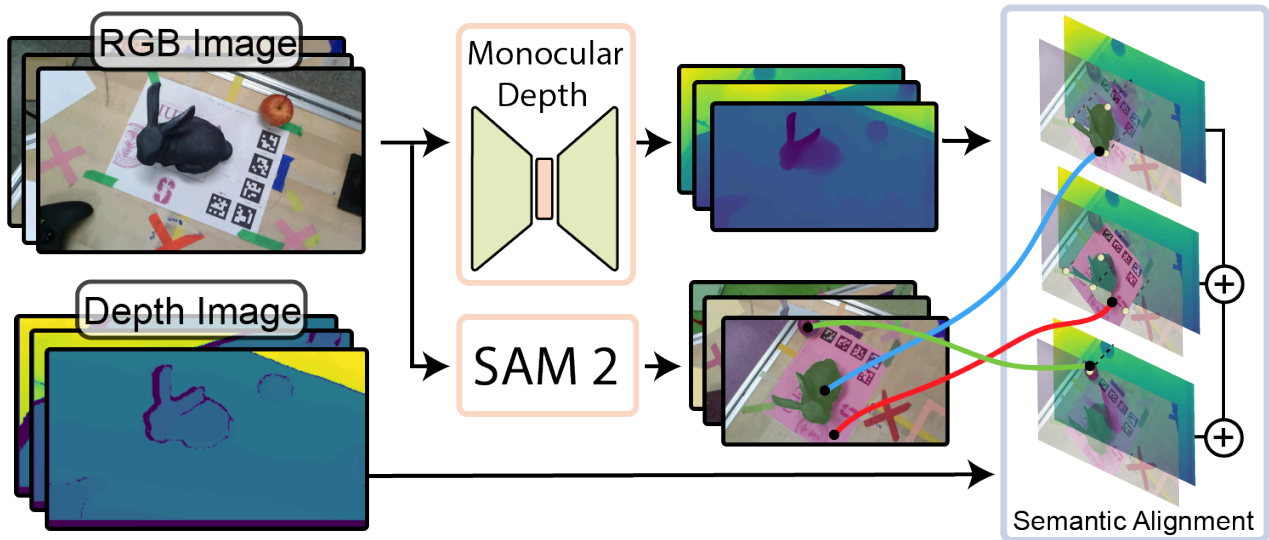


Figure 1: Next Best Sense aligns predicted monocular depth with sensor depth using segmented regions. This version-four figure depicts depth supervision, not the Fisher acquisition rule. Reproduced from [1, Fig. 3; arXiv v4].

NVF instead selects through predicted candidate-image uncertainty, and GO-PRE directs predictive uncertainty reduction toward requested rendered viewpoints [2], [3]. Their geometric or image-quality evaluations do not turn these scores into measured geometric error reduction. Detailed mechanisms and limits are in the NVF and GO-PRE reviews.

Reading status. The detailed review below preserves the earlier literature synthesis and its proposed ARIA-NBV adaptations. Recommendations about model inputs, training stages or implementation are research proposals, not a description of the current system or evidence of measured benefit. The active thesis and implementation specify the current choices.

Active 3DGS and Targeted NBV

Primary sources. ActiveNeRF [4], FisherRF [5], Next Best Sense [1], semantic/dynamic 3DGS NBV [6], instance-aware object-centric 3DGS NBV [7], and FOV-HPE [8].

Source status. The corpus has local TeX mirrors for several 3DGS/NBV papers, including FisherRF, Dynamic 3DGS, Next Best Sense, and Instance/Object-centric NBV. The method-specific review preserves three selected Instance-NBV source figures as PNGs under `figures/external/arXiv/Instance-NBV/`. FOV-HPE is tracked through DOI/PDF evidence, not a local TeX mirror.

Related ARIA-NBV pages. Instance-aware object-centric 3DGS NBV, RRI theory, RL planning, and target-aware thesis questions.

Core contribution

Active NeRF and 3D Gaussian Splatting (3DGS) papers are useful because they separate view utility into uncertainty, information, semantics, object focus, dynamics, or downstream-task loss. They do not replace ARIA-NBV’s utility target. The thesis should keep Relative Reconstruction Improvement (RRI) and Target-Specific RRI (target RRI) as labels/evaluation, and use 3DGS-style signals as proposal or diagnostic channels.

method	verified paper signal	ARIA-NBV adoption	do not adopt
ActiveNeRF [4]	Sparse-view NeRF acquisition uses uncertainty to choose views under a limited budget.	Use uncertainty-like scores as candidate proposal features and rank-agreement diagnostics.	Do not replace mesh-supervised RRI with NeRF uncertainty unless calibrated against RRI.
FisherRF [5]	Fisher information estimates how much a candidate view should reduce radiance-field uncertainty.	Treat Fisher-style information gain as a proposal channel for target-local evidence and uncertainty.	Do not treat Fisher information as GT reconstruction-quality improvement.
Next Best Sense [1]	Robotic 3DGS active sensing combines semantic depth alignment with Fisher-style depth uncertainty for view/touch selection.	Borrow the separation between representation uncertainty, semantic relevance, and downstream diagnostic value.	Do not import touch sensing or 3DGS scene state as thesis-core dependencies.
Semantic/dynamic 3DGS NBV [6]	Active selection can score geometry, semantic Gaussian parameters, and dynamic/deformation parameters separately.	Report scene RRI, target RRI, validity, semantic relevance, and motion cost as separate channels.	Do not make dynamic 3DGS a dependency for static ASE target-conditioned reconstruction.
Instance-aware object-centric 3DGS NBV [7]	Per-Gaussian object logits, opacity, and rendering Jacobians produce block-diagonal RGB/depth/object information gain; a target gate specializes selection to one instance.	Retain EVL as the actor-visible backbone; use damped object-weighted logdet, object entropy, opacity deficit, and target-visible mass as auxiliary candidate channels calibrated against target RRI.	Do not call the analytic rasterizer a learned backbone, substitute logdet for target RRI, expose GT masks/boxes to the actor, or transfer the paper’s greedy sphere-based policy as multi-step planning evidence.
FOV-HPE [8]	DOI/PDF evidence describes dynamic 3DGS novel-view rendering and RL viewpoint refinement for monocular 3D human-pose error.	Treat as evidence that 3DGS can become a downstream-task simulator bridge.	Do not use MPJPE, human-pose reward, or dynamic-human scenes as thesis-core objectives.

Instance-NBV-specific synthesis

The method-specific review distills four points that matter to the thesis:

1. The paper has no learned NBV-policy backbone; its scorer is analytic Gauss–Newton/D-optimal information gain over an object-aware Gaussian map.
2. Its largest depth ablation gain comes from the object-vector reconstruction objective, so reconstruction and view-selection improvements must be separated experimentally.
3. Its target-specific results validate a target/global quality tradeoff, but its policy is greedy and omits path cost, feasibility, stopping, and multi-step value.

- For ATEK snippets, the strongest adaptation is EVL plus semi-dense/ray-aware state with optional Gaussian information channels in candidate tokens, not replacement of the thesis state or reward.

Verified paper signals

The common transferable structure is utility-channel separation:

$$U(q) = \lambda_{\text{geom}} U_{\text{geom}}(q) + \lambda_{\text{target}} U_{\text{target}}(q) + \lambda_{\text{unc}} U_{\text{unc}}(q) + \lambda_{\text{task}} U_{\text{task}}(q).$$

ARIA-NBV should not collapse these into one opaque score. The thesis-core version should log them separately:

ARIA-NBV channel	role
scene RRI	global reconstruction-quality label/evaluation
target RRI	target-conditioned label/evaluation
Validity Mask (validity mask) and invalid reason	feasibility constraint
uncertainty / Fisher / object evidence / observation count	proposal, candidate feature, or diagnostic signal
target/global tradeoff	Pareto analysis under a fixed acquisition budget
motion/path cost	optional planning penalty after quality evidence is trusted

ARIA-NBV adoption

- **Proposal/diagnostic:** uncertainty, Fisher information, object focus, semantic relevance, rank agreement, candidate diversity, target-visible mass, object entropy, and opacity deficit.
- **Gated follow-up:** a persistent object-aware Gaussian map and damped target-weighted logdet only after one-step rank calibration against target RRI.
- **Stretch/bridge:** 3DGS-backed successor simulation, continuous control, or downstream-task simulators after an environment abstraction exists.

Do not adopt

- Do not move the thesis objective from target RRI to 3DGS uncertainty or log-determinant information.
- Do not make 3DGS a required reconstruction backend for ASE oracle rollout generation.
- Do not use GT instance masks, GT OBBs, or GT mesh support in actor-visible candidate scoring.
- Do not present greedy analytic candidate selection as evidence for finite-horizon (Q_H).
- Do not claim local TeX verification for FOV-HPE; the current corpus has DOI/PDF evidence only.

Open risks / caveats

- 3DGS uncertainty and Fisher metrics can disagree with mesh-supervised reconstruction quality.
- Log-determinant rankings depend on damping, parameterization, output scaling, and the block-diagonal approximation.
- Target-aware 3DGS methods often assume segmentation, persistent identity, representation state, and optimization loops that ARIA-NBV does not yet have.
- A pose can be scored without a candidate image, but a leakage-clean rollout successor still requires an acquired, logged, or explicitly simulated observation.
- These papers are most useful as target/proposal inspiration, not as replacements for the finite-candidate Finite-Horizon Q Function (Q_H) thesis path.

Terminology

Metrics

RRI – Relative Reconstruction Improvement: Reduction in reconstruction error relative to the error before one acquisition. Target-specific RRI measures this change on the selected object. Oracle RRI obtains the measurement from privileged reference geometry. Unlike the root-gain reward, its denominator changes with the current reconstruction.

target RRI – Target-Specific RRI: State-relative reconstruction improvement restricted to the selected target. It specializes RRI to the target geometry and differs from the root-normalized reward used to accumulate gain across a rollout.

Model

Q_H – Finite-Horizon Q Function: Expected finite-horizon return after taking a specified candidate action first. The continuation policy determines later choices. The ideal optimal value maximizes over policies using the admitted actor information; a learned scorer approximates it from a finite representation. A sampled return, the fitted training target and the predicted value are distinct. The Q function specifies neither encoder architecture nor training algorithm.

Planning

validity mask – Validity Mask: Boolean indicator whose true entries identify admissible candidate actions. It constrains selection separately from utility prediction. An invalid-action mask has the opposite polarity. Label availability and stored-successor availability require separate indicators.

Representation

3DGS – 3D Gaussian Splatting: Explicit radiance-field representation using optimized 3D Gaussian primitives for real-time novel-view synthesis. ARIA-NBV treats active 3DGS work as proposal, uncertainty, and simulator-bridge literature, not as a replacement for ASE mesh-supervised RRI.

Bibliography

- [1] M. Strong, B. Lei, A. Swann, W. Jiang, K. Daniilidis, and M. K. III, “Next Best Sense: Guiding Vision and Touch with FisherRF for 3D Gaussian Splatting.” [Online]. Available: <https://arxiv.org/abs/2410.04680>
- [2] S. Xue, J. Dill, P. Mathur, F. Dellaert, P. Tsiotras, and D. Xu, “Neural Visibility Field for Uncertainty-Driven Active Mapping,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. [Online]. Available: <https://arxiv.org/abs/2406.06948>
- [3] Y. Song, Z. Li, C. Li, L. He, Y. Wang, and W. Zhang, “GO-PRE: Goal-Oriented Next-Best-View Selection via Predictive Rendering Entropy for Active 3D Reconstruction,” in *Proceedings of the 43rd International Conference on Machine Learning*, in Proceedings of Machine Learning Research, vol. 306. 2026, pp. 114687–114699. [Online]. Available: <https://proceedings.mlr.press/v306/song26k.html>
- [4] X. Pan, Z. Lai, S. Song, and G. Huang, “ActiveNeRF: Learning Where to See with Uncertainty Estimation.” [Online]. Available: <https://arxiv.org/abs/2209.08546>
- [5] W. Jiang, B. Lei, and K. Daniilidis, “FisherRF: Active View Selection and Uncertainty Quantification for Radiance Fields using Fisher Information.” [Online]. Available: <https://arxiv.org/abs/2311.17874>
- [6] Y. Li, W. Jiang, and K. Daniilidis, “Next Best View Selections for Semantic and Dynamic 3D Gaussian Splatting.” [Online]. Available: <https://arxiv.org/abs/2512.22771>
- [7] S. Jeong, E. Lee, J. Kim, and A. Kim, “Informative Object-centric Next Best View for Object-aware 3D Gaussian Splatting in Cluttered Scenes.” [Online]. Available: <https://arxiv.org/abs/2602.08266>
- [8] W. G. Bae, S. Lee, and J. T. Lee, “Finding Optimal Viewpoints for Monocular 3D Human Pose Estimation in Dynamic 3D Gaussian Splatting Space,” in *Proceedings of the IEEE International Conference on Advanced Video and Signal-Based Surveillance*, 2025. doi: 10.1109/AVSS65446.2025.11149906.