Toward a Unified Architecture for Bio-Inspired Active 3D Vision

Oct 17, 2026 · @Luiz Velho

Laboratório VISGRAF · Instituto de Matemática Pura e Aplicada (IMPA)

Technical Report TR-26-__ (number to be assigned) · October 2026

Luiz Velho, IMPA · Claude, Anthropic (collaborator)

Abstract

We propose the *Active Foveal Bayesian* (AFB) architecture, which places the estimator of active Bayesian 3D vision inside the sensing loop of foveal 3D vision.

The two Visgraf research lines on biologically inspired active 3D vision are complementary. Active Bayesian 3D vision (bioeye, active-stereo, bio-3d-vision) built a principled estimator with real stereo matching and first-class uncertainty, but tested it with unlimited uniform sampling, where choosing where to look had little to offer. Foveal 3D vision (fov-3d-vision, f3d-vision) built a narrow, acquiring fovea, persistent head-frame memory and a reproducible 360° baseline, but with oracle correspondence, no uncertainty, and a gaze policy that never reads its own map.

AFB is organised in six layers: a binocular variable-resolution retina, a spherical stereo matcher with three match outcomes, triangulation with propagated variance, a head-frame memory holding a probabilistic surfel map and a direction-state sphere, an oculomotor layer whose vergence is predicted from memory and corrected by out-of-range evidence, and a gaze policy that ranks fixations by availability first, information gain second, and cost. Each design choice is traced to a measured finding of the two lines.

We specify the interfaces between layers, a budget-based evaluation protocol, a six-stage roadmap that changes one component at a time against the frozen f3d-vision baseline, and six pre-registrable hypotheses with their falsifiers. This report is a design proposal: it introduces no new measurements.

Keywords: active vision; foveated sensing; binocular stereo; vergence; Bayesian fusion; gaze control; 360° reconstruction; computational budget.

1 Introduction

A unified architecture is needed because each Visgraf line answered half of the active-vision question, and each half is weakest exactly where the other is strongest.

The companion report, *Bio-Inspired Active 3D Vision: A Comparative Analysis*, reaches three conclusions that motivate this one:

1. **Line (1) has the estimator but not the regime.** Its thirteen experiments found that, under unlimited uniform sampling, the choice of gaze barely changes what a stereo estimator recovers. Availability of measurements dominated their selection by 65:1 on median error, and foveal weighting was strictly lossy.
2. **Line (2) has the regime but not the estimator.** Its greedy explorer reconstructed 98.32% of a 360° room within 12 mm because a narrow fovea genuinely acquires new data at every saccade.

But correspondence came from an oracle, the map carried no uncertainty, and the policy never read the map.

3. **The two halves of line (2) pull in different directions.** fov-3d-vision showed that foveation wins on fixated targets and loses on whole-sphere coverage at equal rays; f3d-vision optimises whole-sphere coverage.

The architecture proposed here resolves all three. It keeps line (2)'s loop, sensing regime, interfaces and reproducibility discipline; inserts line (1)'s real matching, uncertainty and vergence–scale coupling; and gives the fovea and the periphery different jobs so that foveation and coverage stop competing.

1.1 What this report specifies

- Seven design principles, each traced to the finding it answers (Section 2).
- The layers of the architecture and the data that crosses each interface (Sections 3 and 4).
- The gaze policy in closed form (Section 5).
- An evaluation protocol built on a declared computational budget (Section 6).
- A staged roadmap and pre-registrable hypotheses (Section 7).
- Risks, limitations and open questions (Section 8).

1.2 What it does not claim

Nothing here has been run. The architecture is a proposal whose parts are justified by measured findings, not a result. The design keeps the fixed head of both lines until its last stage, assumes static scenes, and uses no learned components; Section 8 returns to these limits.

2 Design principles

Seven principles fix the architecture, and each one answers a specific measurement rather than an intuition.

Principle	Evidence it answers	Consequence in the design
P1. Foveation is sampling under a declared budget, not weighting.	Foveal weighting on a uniform sensor was strictly lossy on every seed (bio-3d exp013). fov-3d's camera renders 50.8× fewer rays than uniform sampling.	A variable-resolution retina (L0). Every comparison is made at equal budget against uniform sampling.
P2. The periphery is coarse, not discarded.	Uniform sampling beat foveation over the whole sphere at all but the largest budget (fov-3d A5). f3d renders a 29.4° field and uses only its 12° core.	Peripheral samples drive coverage, coarse depth, vergence presets and saccade targets. The fovea is reserved for precision.
P3. Availability before selection.	Knowing where measurement is possible beat knowing how uncertain it is by 65:1 (exp002). Closing the loop paid off as coverage over accuracy by ≥33× (exp011).	The policy's primary term is availability. Information gain is a secondary term whose weight is measured, not assumed.

Principle	Evidence it answers	Consequence in the design
P4. Every measurement has three outcomes.	A vergence controller driven by valid disparities verged away from its target, because the evidence it needed had failed validity (exp003).	The matcher reports *matched*, *occluded* or *out of range* (near or far). Out-of-range evidence feeds vergence and memory.
P5. Vergence stays in the loop; metric depth comes from exact triangulation.	Vergence sets the metric scale of disparity (framework scaling closure). Linearizing about fixation filters confident-wrong matches (TR-26-09). f3d triangulates exactly on the sphere.	Vergence sets the disparity search band. The linearization offset only inflates variance; it never computes depth.
P6. The sensor must be able to acquire.	On a complete spherical sensor a saccade returns an identical capture (bio-3d, measured against Blender). f3d's narrow field acquires at every saccade.	A narrow, variable-resolution field. Head translation is the action reserved for surfaces that no eye rotation can reach.
P7. Separable, reproducible and falsifiable.	f3d's truth-stripped interfaces and bitwise baseline. bio-3d's ledgers, pre-registration and characterised stimuli.	Narrow interfaces per layer, an oracle arm kept permanently, a frozen baseline, and pre-registered falsifiers for every stage.

Two principles carry most of the weight. P2 removes the tension inside line (2), because the periphery covers the sphere cheaply while the fovea is spent where precision matters. P4 removes the most general failure mode found in line (1), and it has a biological counterpart: outside Panum's fusional band, binocular vision does not go silent but reports diplopia, itself a coarse depth signal.

3 Architecture overview

AFB is one loop of six layers in which perception writes to a single head-frame memory and action reads from it.

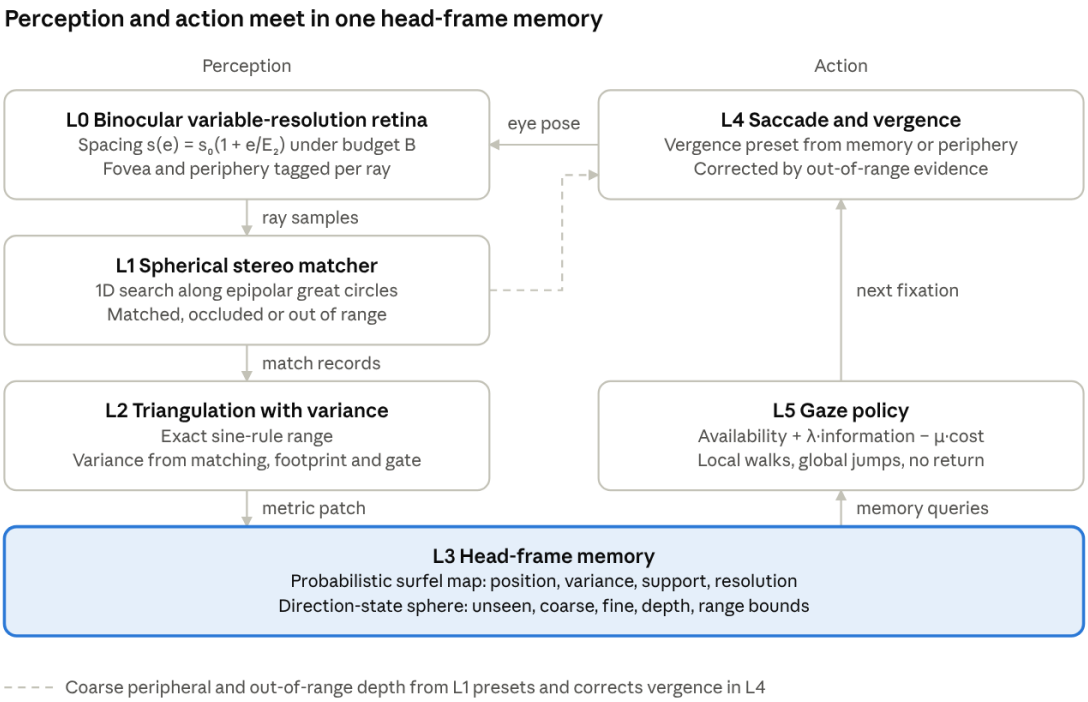


Figure 1 · the AFB architecture: six layers, one memory, two loops

The perception path (L0–L2) turns a fixation into a metric patch with variance. The action path (L5, L4) reads the memory to choose the next gaze and its vergence. A second, faster link carries coarse and out-of-range depth straight from the matcher to vergence, so that the eyes converge on what they are about to measure.

3.1 Where each component comes from

Layer	Taken from	What changes
L0 Retina	fov-3d-vision's foveated camera; w3d-scenes' fixed eye	Made binocular; the periphery is kept and used
L1 Matcher	bio-3d-vision's matcher and validity tests; f3d-vision's correspondence interface	Runs on a spherical epipolar lattice; adds the out-of-range outcome
L2 Local metric	f3d-vision's exact triangulation; bioeye's Jacobian; bio-3d-vision's linearization gate	Variance propagated for every point
L3 Memory	f3d-vision's surfel map and coverage sphere; bioeye's precision fusion; fov-3d-vision's finest-owns rule	Variance and resolution per surfel; coarse and bounded cell states
L4 Oculomotor	The framework's vergence layer; bio-3d-vision's Listing's-law geometry	Vergence preset from memory and corrected by out-of-range evidence (new)
L5 Policy	f3d-vision's local and global saccades; bioeye's uncertainty-driven gaze	Availability-first value, map frontiers, travel-ordered jumps (new)

3.2 Two nested loops

The framework's nesting of a fast inner loop inside a slow outer loop survives in a new form. The inner loop is vergence within one fixation: preset, measure, correct. The outer loop is gaze across fixations: write to memory, read from memory, saccade. Both close through the same memory, which is the reference-frame closure of the framework made concrete.

4 Layer specifications

Each layer is defined by what it receives and what it emits, so that any one can be replaced while the others stay fixed. Frames and conventions follow f3d-vision: the fixed head frame H0 has its origin between the eyes, +X along the baseline, +Y up and −Z forward; lengths are in metres.

4.1 Interfaces

Interface	From → to	Contents
Eye pose	L4 → L0	Gaze (ψ, θ) in H0, vergence distance D, torsion from the Listing tilt parameter
Ray samples	L0 → L1	Per eye: sample direction in H0, angular footprint ω, radiance; fovea and periphery tagged
Match record	L1 → L2	Left and right sample indices, angular disparity η, its variance $\sigma\eta^2$, status: matched, occluded, out-of-range-near, out-of-range-far
Metric patch	L2 → L3	Points in H0 with range variance and footprint; range bounds for out-of-range rays
Memory queries	L3 → L4, L5	Predicted range and variance along any direction; per-cell state, resolution and variance on the sphere
Fixation command	L5 → L4	Next gaze and its expected value; local or global saccade

The match record extends f3d-vision's truth-stripped correspondence interface. The oracle keeps emitting it, with status taken from its visibility test, so the oracle remains a permanent comparison arm.

4.2 L0 — Binocular variable-resolution retina

L0 renders, for each eye, ray samples whose angular spacing grows with eccentricity from the eye's gaze. It is fov-3d-vision's foveated Cycles camera made binocular.

- **Geometry.** Eye centres at $(\mp b/2, 0, 0)$ in H0 with b = 63 mm, rotating about their centres. Torsion follows Listing's law with a tilt parameter; bio-3d's default is 0.25 per radian of vergence, and 0.5 cancels vertical disparity in the plane of regard.
- **Sampling law.** Spacing grows linearly with eccentricity e, the cortical-magnification form used to size fov-3d's calibration targets, up to a maximum eccentricity e_max:

$$s(e) = s_0 \left(1 + \frac{e}{E_2}\right), \qquad 0 \le e \le e_{\max}$$

- **Parameters.** fov-3d's settled values are $s_0 = 0.1°$ (small profile) or 0.05° (full), $E_2 = 2°$ and e_max = 45°, with E_2 left to the objective (D11).
- **Budget accounting.** The number of rays per fixation is a function of (s_0, E_2, e_max) and is charged to the run's budget (Section 6).

4.3 L1 — Spherical stereo matcher

L1 matches samples along epipolar great circles and reports one of three outcomes for every foveal sample.

- **Epipolar coordinates.** For a direction d in H0, ϑ is its angle from the baseline and ϕ = atan2(d_y, −d_z) its rotation about it. Corresponding rays share ϕ, so matching is a one-dimensional search in ϑ. With both eye centres fixed, this geometry is constant in the head frame.

- **Lattice.** Interim: both eyes' samples are resampled onto a shared (ϑ, ϕ) lattice whose spacing follows s(e). Target: a matcher that operates natively on the non-uniform lattice, scoped as research.
- **Search band.** Centred on the disparity predicted by the current vergence, with a width that grows with eccentricity, in the manner of Panum's fusional area.
- **Coarse to fine.** Coarse peripheral matches initialise and bound the fine foveal search, as in the multiresolution layer of the framework.
- **Outcomes.** *Matched*, with η and ση² from the curvature of the cost curve and the best-to-second-best ratio. *Occluded*, when left–right consistency fails inside the band. *Out of range*, near or far, when the cost minimum sits on the band's edge.

4.4 L2 — Triangulation with variance

L2 turns matches into metric points with a variance, and out-of-range rays into range bounds. Range from the left eye is exact, by the sine rule used in both bio-3d-vision and f3d-vision, with η = ϑ_R − ϑ_L:

$$r_L = \frac{b \sin \vartheta_R}{\sin \eta}, \qquad \frac{\partial r_L}{\partial \eta} = -\frac{b \sin \vartheta_L}{\sin^2 \eta}$$

For small η the derivative reduces to −r²/(b sin ϑ_L), which at ϑ_L = 90° is bioeye's Jacobian −D²/I. Range variance combines three terms: matching noise, the sample footprint at that eccentricity, and a gate that grows with the square of the offset from the fixated disparity η_fix:

$$\sigma_r^2 = \left(\frac{b \sin \vartheta_L}{\sin^2 \eta}\right)^2 \left(\sigma_\eta^2 + k\,\omega^2\right) + c\,(\eta - \eta_{\mathrm{fix}})^2$$

The constants k and c are parameters to be calibrated. The gate term is the confidence mechanism identified in TR-26-09: it down-weights estimates far from the fixated depth, where confident-wrong matches concentrate. An out-of-range-near ray contributes the bound r < r_near along its direction; out-of-range-far contributes r > r_far.

4.5 L3 — Head-frame memory

L3 keeps two linked structures in H0: a probabilistic surfel map for geometry, and a direction-state sphere for attention.

- **Surfels.** Each holds position (float64), range variance σ² along its ray, support count, the finest footprint that measured it, first fixation and provenance. Colour is kept for display only, as in f3d.
- **Association.** Nearest surfel within max(12 mm, 3σ), against a snapshot of the map taken before the patch, as in f3d's exact 27-cell hash search.
- **Fusion.** Precision-weighted, with a Mahalanobis outlier gate. When two resolutions differ by more than a set factor, the finer measurement owns the surfel, following fov-3d's finest-owns rule.

$$x' = \frac{x/\sigma^2 + \bar{x}/\tau^2}{1/\sigma^2 + 1/\tau^2}, \qquad \sigma'^2 = \left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)^{-1}$$

- **Direction-state sphere.** A 1° grid, as in f3d, or an equal-area grid. Each cell records a monotone state — unseen, coarse, fine, depth — plus the best resolution achieved, the smallest depth variance, and any near or far bounds. All fractions are solid-angle weighted.

- **Remapping.** Because everything lives in H0, trans-saccadic remapping is the identity; the efference copy is the known eye rotation.

4.6 L4 — Saccade and vergence

L4 points both eyes and chooses a vergence distance before the measurement, then corrects it from the measurement's own evidence.

1. **Preset.** At the target direction, take the predicted range from the map if the cell holds depth; otherwise the coarse peripheral estimate; otherwise f3d's default of 2.1 m.
2. **Measure** with the band centred on the preset.
3. **Correct.** If out-of-range-near outcomes dominate the fovea, converge; if out-of-range-far dominate, diverge; otherwise set vergence from the robust median of matched foveal disparities. At most one corrective re-verge per fixation, charged to the budget.

This replaces f3d's fixed convergence point and removes the failure of bio-3d exp003, in which the controller could not see targets outside its band.

4.7 L5 — Gaze policy

L5 reads L3 and the current fixation's match record and returns the next fixation; Section 5 specifies it.

5 Gaze policy

The policy ranks candidate fixations by availability first, information gain second, and cost, and keeps f3d-vision's two-tier structure of local walks and global jumps.

5.1 Value of a fixation

For a candidate gaze g with foveal footprint F(g) on the direction-state sphere:

$$V(g) = A(g) + \lambda\, I(g) - \mu\, C(g)$$

- **Availability A(g)** is the solid angle in F(g) that the fovea would newly resolve. A cell contributes its solid angle $\Delta\Omega$ times its resolution deficit: 1 if unseen, and the fraction by which its best footprint exceeds the fovea's if it was seen only coarsely.
- **Information gain I(g)** is the expected reduction of depth variance over cells in F(g) that already hold depth. For a Gaussian cell of variance σ^2 and a predicted measurement variance τ^2 at that range and eccentricity (from L2), the closed form is:

$$I(g) = \sum_{c \in F(g)} \Delta\Omega_c \, \frac{\sigma_c^4}{\sigma_c^2 + \tau_c^2}$$

- **Cost C(g)** is the saccade amplitude plus the rays and matcher operations the fixation will consume.
- **Weights.** λ and μ are measured, not assumed. Line (1) predicts that the best λ is small: availability should do most of the work, with information gain as a second-order refinement.

5.2 Local and global saccades

- **Local walk.** As in f3d, eight neighbours at 7.2° in the current tangent frame. f3d's `edge_support` term is kept, but computed so that occluded and out-of-range samples count as *possible surface*,

not as absence. Without that change, a natural matcher would make unmatched surfaces look empty and cut local walks short (Section 6.4 of the companion report).

- **Global jump.** Candidates are the centres of unseen components, as in f3d, *plus* map frontiers: depth discontinuities, low-support surfels and surfaces seen only at grazing angles. Candidates are ordered by V(g) per unit of travel, not by size alone. Several small holes may be covered by one fixation when they fall in the same footprint.
- **Inhibition of return.** No fixation within 2° of a visited gaze, as in f3d, unless its information gain exceeds a threshold. This prevents the lockup that bio-3d's maximum-variance loop showed.

5.3 Peripheral preview

Coarse peripheral samples are matched at every fixation. They update the direction-state sphere to *coarse*, provide range estimates for L4's vergence preset, and seed global candidates before the fovea arrives. This is the biological division of labour between a coarse periphery that decides where to look and a fovea that resolves what is there.

5.4 Stopping

The loop stops when the declared budget is spent, when 99% of the sphere is seen and the largest depth variance in fine cells is below a target, or when no eligible candidate remains. The 600-fixation safety cap of f3d is kept.

6 Evaluation protocol

The architecture is judged by map quality as a function of a declared computational budget, which is the claim fov-3d-vision set out to test: the same posterior for less computation.

6.1 The budget

Every run declares its budget B before it starts. B counts three hardware-independent quantities: rays rendered, matcher hypotheses evaluated, and fixations, each re-verge included. Wall time is recorded for every stage but is not part of B, because it depends on the machine.

6.2 Metrics

Family	Metric	Origin
Coverage	Fraction of first-hit reference solid angle with a surfel within 12 mm	f3d-vision
Accuracy	Median and 90th-percentile distance to the reference, reported separately within and away from depth discontinuities	f3d-vision; banding from bio-3d-vision
Calibration	Negative log-likelihood and empirical coverage of 2σ intervals of surfel range	bioeye validation agenda
Efficiency	Fixations, rays and matcher hypotheses to reach 99% seen; length of the hole-filling tail	f3d-vision; fov-3d-vision
Footprint	Surfels and bytes in the final map	f3d-vision

Results are reported as curves of each metric against B, never as a single operating point.

6.3 Comparison arms

Every stage is compared against fixed references, so that a difference can only come from the component that changed:

- **Official baseline.** f3d-vision's tagged Greedy Foveal Explorer, reproduced bitwise.
- **Oracle arm.** The same architecture with oracle correspondence through the same match record: the ceiling for matching.
- **Uniform arm.** The same architecture with uniform sampling at equal B: the control for foveation.
- **Policy arms.** $\lambda = 0$ against $\lambda > 0$; coverage-only against target-driven global jumps.

6.4 Stimuli

- **Calibration room** from w3d-scenes: known targets at known eccentricities and distances, for checking the retina and the variance model.
- **Classroom:** the scene of f3d's baseline, for continuity with its numbers.
- **HDRI tier:** an environment at infinity, for monocular checks of sampling against its own ground truth.
- **A contrast-controlled occlusion stimulus,** proposed in TR-26-09, to separate the effect of contrast from that of realism at occlusion boundaries.

Each stimulus is characterised before it is used to measure, following TR-26-09: the oracle's behaviour at occlusion boundaries in particular is measured, not assumed.

6.5 Discipline

The protocol combines the practices of both lines:

- From f3d-vision: a frozen, tagged baseline; bitwise reproduction of every control-path product; a truth firewall that keeps the reference closed until a run is frozen.
- From bio-3d-vision: a pre-registration for every experiment stating both what would falsify the hypothesis and what would mean the experiment asked the wrong question; a decision ledger and a measurement ledger; several seeds, with spread reported.

7 Roadmap and hypotheses

The architecture is reached in six stages from the f3d-vision baseline, each changing one component and each closed by a pre-registered experiment before the next begins.

7.1 Stages

Stage	The one component changed	Primary measurement	Gate to the next stage
0	None: f3d-vision official baseline	Already measured: 542 fixations, 98.32% within 12 mm	Done
1	Natural matcher on the uniform 12° fovea, emitting the three-outcome match record	Coverage and accuracy against the oracle arm, by band	The loop still reaches its stopping rule; losses localised and explained
2	Variance in the map, precision fusion, and the information-gain term in the policy	Discontinuity error at equal fixations; calibration of 2σ intervals	Intervals calibrated within a pre-set tolerance
3	Peripheral preview, vergence preset and out-of-range correction	Fixations to 99% seen; length of the hole-filling tail; vergence error at targets	Vergence converges on targets outside the initial band
4	Binocular variable-resolution retina (fov-3d Phase B): interim resampling, then a native non-uniform matcher	Rays per fixation; coverage and accuracy at equal B	Matching on the foveated lattice is no worse than on the uniform one in the fovea
5	None in the architecture: budget sweeps across arms	Quality-against-budget curves, foveated against uniform	—
6	Head translation as an action	Coverage of surfaces hidden from the head origin	—

Stages 1–3 do not depend on the foveated retina, so the research risk of Stage 4 does not block them.

7.2 Hypotheses

Each hypothesis is stated with the result that would falsify it, ready for pre-registration.

Id	Stage	Hypothesis	Falsified if
H1	1	Natural matching costs coverage mainly at depth discontinuities and in textureless or see-through regions, not uniformly	The loss relative to the oracle is the same in every band
H2	1	Counting occluded and out-of-range samples as possible surface in `edge_support` restores the length of local walks lost to natural matching	The ratio of local to global saccades is unchanged by the fix
H3	2	The information-gain term lowers discontinuity error at equal fixations, but changes fixations-to-99% far less than availability does	Information gain moves fixations-to-99% more than availability, or has no effect on error
H4	3	Peripheral preview plus travel-ordered global jumps shortens the hole-filling tail (92 fixations in the baseline) and the total to 99%	Neither count falls
H5	3	A vergence preset from memory and periphery, corrected by out-of-range evidence, removes the verge-away failure of bio-3d exp003	Vergence still diverges from targets that start outside the band
H6	5	At equal B, foveated sensing under a target-driven policy matches uniform sensing in coverage and beats it at discontinuities; under a coverage-only policy it does not	Uniform sampling is at least as good under both policies

H6 is the central claim. Its second clause deliberately reproduces fov-3d-vision's Phase A result as a control: if foveation suddenly beat uniform sampling under the coverage-only policy too, the measurement would be suspect.

8 Risks and limitations

The largest risk is the matcher on a non-uniform lattice; the roadmap is ordered so that everything else can be validated before it lands.

Risk	Consequence	Mitigation
A native non-uniform matcher is a research problem, not engineering	Stage 4 may take far longer than the others	Interim resampling to a shared epipolar lattice; Stages 1–3 are independent of it
Natural matching cuts coverage sharply on textureless walls and windows	Absolute numbers fall well below the baseline's	Report every result relative to the oracle arm; coarse-to-fine matching; texture-rich stimuli alongside the Classroom
Variance is miscalibrated	The policy chases artefacts, as bio-3d's maximum-variance loop did	Calibration gate at Stage 2; inhibition of return; availability as the primary term

Risk	Consequence	Mitigation
Rendering dominates cost	In f3d, rendering took 413.5 of 733 s; in bio-3d, renders were 91% of the closed loop's cost	The foveated retina cuts rays; the persistent render server is kept
The map grows without bound	f3d's map reached about 1 GB, and fusion time grew from 0.23 to 0.79 s per fixation	Compact surfels and finest-owns pruning, off the critical path
Results stay synthetic	Conclusions may not transfer to real imagery	Characterised stimuli now; real calibrated stereo as a later stimulus

8.1 Limitations of the proposal

- The head is fixed until Stage 6, so surfaces hidden from the head origin remain outside the target, as in f3d.
- Scenes are static, and no learning or semantic component is included.
- The information-gain term may prove second-order, as line (1) predicts. That outcome would still confirm, on a far stronger testbed, that availability dominates selection.

9 Conclusion

AFB keeps what each line got right and replaces what each got wrong. From foveal 3D vision it takes a narrow fovea that acquires at every saccade, a persistent head-frame memory, a separable loop and a reproducible baseline. From active Bayesian 3D vision it takes real matching, propagated uncertainty, the coupling of vergence and metric scale, and a catalogue of failure modes to design against. Its two new ideas are small: give the coarse periphery the job of coverage so that foveation and coverage stop competing, and let every measurement say when it is out of range. Whether those ideas deliver the same map for less computation is the question the roadmap is built to answer.

References

1. J. Aloimonos, I. Weiss and A. Bandyopadhyay. Active vision. *International Journal of Computer Vision* 1(4):333–356, 1988.
2. R. Bajcsy. Active perception. *Proceedings of the IEEE* 76(8), 1988.
3. D. H. Ballard. Animate vision. *Artificial Intelligence* 48(1):57–86, 1991.
4. K. Friston, F. Rigoli, D. Ognibene, C. Mathys, T. FitzGerald and G. Pezzulo. Active inference and epistemic value. *Cognitive Neuroscience* 6:187–214, 2015.
5. D. Marr and T. Poggio. A computational theory of human stereo vision. *Proc. R. Soc. Lond. B* 204:301–328, 1979.
6. J. Najemnik and W. S. Geisler. Optimal eye movement strategies in visual search. *Nature* 434:387–391, 2005.
7. E. L. Schwartz, D. N. Greve and G. Bonmassar. Space-variant active vision: definition, overview and examples. *Neural Networks* 8(7–8):1297–1308, 1995.

8. A. L. Yarbus. *Eye Movements and Vision*. Plenum Press, 1967.
9. L. Velho (with GPT). A framework for stereoscopic 3D vision: active Bayesian inference with nested oculomotor control. Working draft, VISGRAF/IMPA, August 2026.
10. L. Velho (with GPT). Bioeye: a bio-inspired active Bayesian stereo architecture. Technical report, VISGRAF/IMPA, September 2026.
11. L. Velho and Claude. Experiments on active 3D vision. Technical Report TR-26-09, VISGRAF/IMPA, 2026.
12. L. Velho, ChatGPT and Claude Code. Active foveal 3-D vision: a fixate–reconstruct–saccade playground for 360-degree scene reconstruction. Technical Report TR-26-11, VISGRAF/IMPA, October 2026.
13. L. Velho and Claude. Bio-inspired active 3D vision: a comparative analysis of active Bayesian and foveal approaches. Technical report, VISGRAF/IMPA, October 2026 (companion report).

Repositories

- visgraf/bio-3d-vision
- visgraf/fov-3d-vision
- visgraf/f3d-vision, tag `greedy-foveal-playground-baseline`
- visgraf/bioeye
- visgraf/w3d-scenes (archived)

Appendix A: Provenance

This report was drafted by Claude (Anthropic, Opus 5.5) at the request of L. Velho, who set its scope; it develops an architecture first sketched in conversation on 10 October 2026. It is a design proposal and contains no new measurements. Numbers quoted as evidence come from the technical reports and repository README files listed above, read on 10 and 17 October 2026. Formulas in Section 4 restate relations from those sources, except the variance model of Section 4.4 and the policy of Section 5, which are proposed here and remain to be calibrated.