Bio-Inspired Active 3D Vision: A Comparative Analysis of Active Bayesian and Foveal Approaches

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

Two Visgraf research lines pursue biologically inspired active 3D vision, and they differ less in geometry than in what the eye movement is for.

The first line, *active Bayesian 3D vision* (bioeye, active-stereo, bio-3d-vision), treats gaze as a control on a depth estimator. Fixation and vergence set the metric scale of stereo, and the next fixation goes where the depth posterior is least certain. The second line, *foveal 3D vision* (fov-3d-vision, f3d-vision), treats gaze as the way a narrow, high-quality fovea and a persistent memory assemble a wide-field 3D percept.

The first line's thirteen pre-registered experiments gave a consistent negative answer to its own question. Availability of measurements dominated their selection, foveal weighting on a uniform sensor was strictly lossy, and the one positive result was a coverage effect. The second line built its system in exactly that regime: a greedy explorer with a 12° binocular fovea and oracle correspondence reconstructed 98.32% of the first-hit solid angle of a 360° room within 12 mm, in 542 fixations, reproducible bit for bit.

We compare the two lines' architectures, evidence, strengths and limitations. We argue that they are complementary halves of one system. The first contributes the estimator, first-class uncertainty and the failure modes of real stereo matching. The second contributes a sensing regime in which looking acquires new data, a separable architecture, and the beginnings of budgeted foveated sampling. A companion report proposes an architecture that unifies them.

Keywords: active vision; binocular stereo; foveation; vergence; Bayesian fusion; saccadic exploration; 360° reconstruction.

1 Introduction

This report compares two Visgraf approaches to biologically inspired active 3D vision and finds that they answer different questions about the same loop.

The active-perception tradition holds that perception is better posed as controlled sampling than as passive reconstruction [Bajcsy 1988; Aloimonos et al. 1988; Ballard 1991]. Both lines studied here take that claim seriously. Both are binocular, keep the head fixed, rotate the eyes about their own centres, and build a percept over a sequence of fixations. They were framed from the outset as two readings of the same idea:

1. **Active Bayesian 3D vision** poses 3D vision as a *controlled sampling problem*. Its question is whether choosing where to look improves binocular depth estimation.

2. **Foveal 3D vision** poses it as a *reconstruction problem*. Its question is whether a spatially selective binocular observer can build a persistent metric representation of its whole surroundings by deciding where to look.

The comparison shows that the sharper distinction is between *estimation* and *availability*. Line (1) uses gaze to improve an estimate it already holds; line (2) uses gaze to acquire measurements it does not yet have. That distinction explains most of the differences in architecture, and most of the differences in outcome.

1.1 Scope and corpus

The analysis rests on the project's own documents and the public state of its repositories as of 17 October 2026:

- *A Framework for Stereoscopic 3D Vision: Active Bayesian Inference with Nested Oculomotor Control*, working draft, August 2026.
- *Bioeye: A Bio-Inspired Active Bayesian Stereo Architecture*, technical report, September 2026.
- *Experiments on Active 3D Vision*, TR-26-09, September 2026.
- *Active Foveal 3-D Vision: A Fixate–Reconstruct–Saccade Playground*, TR-26-11, October 2026.
- The README files of bioeye, bio-3d-vision, fov-3d-vision, f3d-vision and w3d-scenes.

The active-stereo repository was not publicly reachable when this report was prepared; it is described through the records its successors keep of it. For f3d-vision the official, tagged baseline is used throughout. Playground variants accepted after it and kept on their own branches are out of scope. No measurement was repeated for this report: every number is quoted from the source that produced it.

1.2 Outline

Section 2 introduces the two lines and their repositories. Section 3 compares their architectures. Section 4 summarises what each line measured, and Section 5 sets out strengths and weaknesses. Section 6 discusses how the lines relate, and Section 7 concludes.

2 The two research lines

Six repositories carry the work: three for line (1), two for line (2), and one shared synthetic eye that is now part of line (2).

Repository	Line	Role	Sensor and geometry	State, October 2026
bioeye	1	Thin vertical slice: one script runs L0–L5 end to end	Rectified parallel pinhole, scalar vergence	Demonstration; 432-line script, no tests

Repository	Line	Role	Sensor and geometry	State, October 2026
active-stereo	1	Full implementation with verified oculomotor geometry, a decision record and a physically based rendering path	Converged eyes under Listing's law	The active loop never ran; superseded
bio-3d-vision	1	Experiment repository: the running loop of bioeye with the rigour of active-stereo	Pinhole and spherical sampling; head-anchored belief	Closed: 13 pre-registered experiments, report TR-26-09
fov-3d-vision	2	Variable-resolution rendering for machine vision: the sampling side	Foveated OSL camera in Cycles; monocular	Phase A closed pending D11; Phase B (binocular) not started
f3d-vision	2	Active foveal playground: the Greedy Foveal Explorer	Binocular, 12° core of a uniform perspective render, spherical triangulation	Official baseline tagged and reproduced bitwise; report TR-26-11
w3d-scenes	Shared	Synthetic 360° eye and test scenes in three tiers	Equirectangular view from a fixed eye	Archived 12 September 2026; folded into fov-3d-vision as step A1

2.1 Line (1): active Bayesian 3D vision

Line (1) grew from a theoretical framework into three successive implementations, each built to fix what the last one could not do.

The framework models stereoscopic vision as one recursive Bayesian estimator with a control policy. It is organised in six layers: generative geometry, neural disparity encoding, disparity-field inference, metric scaling and cue fusion, oculomotor control, and an active-sampling outer loop. Three closures bind them: disparity is metrically inert without a distance estimate from vergence; fast vergence feedback and slow gaze policy share one estimation-and-control language; and eye posture enters both the forward model and the trans-saccadic estimator.

The implementations instantiate that framework in stages:

- **bioeye** closed the *accumulation* loop. A per-pixel Gaussian posterior is fused across fixations under a greedy maximum-variance policy. It did not close the *perception–action* loop: the disparity field was computed once and re-weighted at each fixation.
- **active-stereo** built the full machinery, but layer by layer the active components received a fraction of the code the input side did, and no experiment exercised them.
- **bio-3d-vision** combined a running loop with a decision ledger, a measurement ledger and pre-registered falsifiers. Its thirteen experiments are the evidence base of Section 4.

2.2 Line (2): foveal 3D vision

Line (2) splits the problem along a different seam: one repository for the sensor, one for the loop.

- **fov-3d-vision** asks what gaze-directed variable-resolution sampling costs and what it buys. Its stated target is *the same posterior for less computation*, under a declared budget. Phase A built a foveated Cycles camera, sequences of fixations, and integration into a spherical representation, all monocular.
- **f3d-vision** asks whether a narrow binocular fovea, moved by saccades and backed by persistent memory, can reconstruct a whole 360° scene. Its official baseline runs one loop: fixate, reconstruct everything the fovea can match, update a global map and a coverage sphere, saccade. Stereo correspondence is supplied by an oracle derived from the renderer, behind an interface a natural matcher can replace.

2.3 The shared synthetic eye

Both lines ended up rendering the same kind of world. w3d-scenes verified three scene tiers from a fixed eye position: an HDRI environment, a generated calibration room, and Blender's Classroom. Its Classroom eye sits at (−0.6, −1.0, 1.2) m, the same point f3d-vision uses as its fixed head origin. bio-3d-vision used Blender renders as its second stimulus alongside an analytic random-dot fixture.

3 Architectures compared

Both lines run the same loop shape — fixate, measure, remember, move — but differ at every stage in what is computed and, above all, in what drives the next fixation.

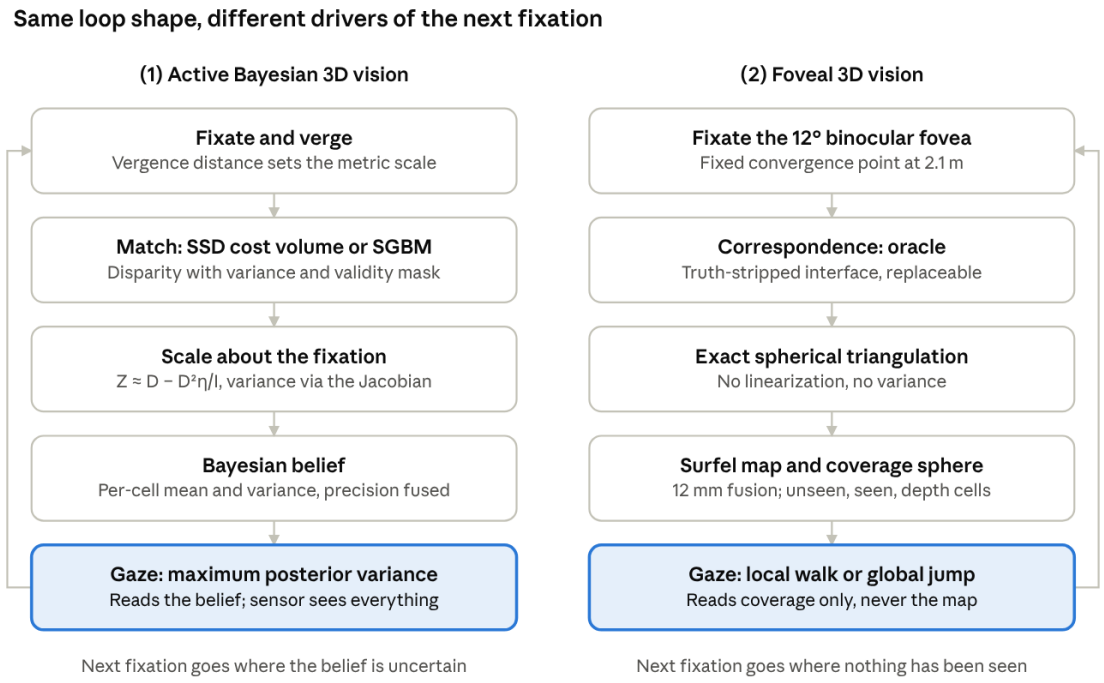


Figure 1 · the two fixation loops, stage by stage

In line (1) each stage refines a probabilistic estimate, and the gaze policy reads that estimate. In line (2) each stage adds geometry to a persistent map, and the gaze policy reads only a record of where the eyes have already looked.

3.1 Common ground

The two lines share more geometry than their framings suggest:

- A fixed head with two eyes rotating about their own centres, separated by a 63–65 mm baseline.
- Spherical epipolar geometry. With the baseline along +x, epipolar great circles are the meridians about it, and range follows from the sine rule $r_l = b \sin \theta_r / \sin(\theta_r - \theta_l)$. bio-3d-vision's `sampling.py` and f3d-vision's triangulation use the same relation.
- A head-anchored memory that stores directions rather than pixels, so it survives every eye movement unchanged.
- Blender Cycles as the source of photorealistic stimuli; the Classroom scene is shared by fov-3d-vision and f3d-vision.

3.2 Stage-by-stage differences

Stage	(1) Active Bayesian	(2) Foveal
Persistent state	Gaussian belief: per-pixel mean and variance (bioeye); per-direction, head-anchored (bio-3d)	Surfel map M (position, colour, support, first fixation), coverage sphere S at 1°, visited gazes V
Measurement	Real matcher: SSD cost volume or SGBM, subpixel decoding, left–right consistency	Oracle correspondence returning four arrays; geometry stripped from its output
Role of vergence	Part of the measurement model: d_fix gives D, which sets the Jacobian and the metric depth	Geometric only: both eyes converge on a point 2.1 m along the gaze
Oculomotor model	Listing's law with a tilt parameter; vergence controller	Eyes rotate about fixed centres; image baseline kept horizontal; no torsion model
Metric depth	Linearized about fixation, $Z \approx D - (D^2/I)\eta$, with variance $\propto D^4$	Exact triangulation, numerically guarded; no variance
Fusion	Precision-weighted Gaussian update	Support-weighted mean within a 12 mm association radius
Policy input	The belief: maximum variance, or expected variance reduction	Coverage sphere, visited gazes and the current fovea's valid-geometry mask
Foveation	A gaze-dependent weight on uniformly sampled data	f3d: the central 12° of each render, periphery unused; fov-3d: a variable-resolution warp
Field of view	Pinhole (bioeye), then a complete sphere (bio-3d)	29.4° per eye, of which a 12° core is measured
Stopping	A fixed fixation budget	99% of the sphere seen, with a cap of 600 fixations
Evaluation	Depth error against ground truth, policy ablations, pre-registered falsifiers	Coverage of a first-hit reference within 12 mm, behind a truth firewall

3.3 Layer correspondence

The framework's layer vocabulary maps onto line (2)'s services only partially. Geometry (L0) and metric scaling (L3) correspond to f3d-vision's sensor geometry and triangulation. Disparity encoding and inference (L1–L2) are replaced wholesale by the oracle. Oculomotor control (L4) is reduced to pointing both eyes at a gaze. The active outer loop (L5) exists in both, but in line (2) it optimises coverage rather than information. Line (2) adds what the framework leaves implicit: an explicit global map, a coverage memory, and a stopping rule.

4 Evidence: what each line measured

Line (1) measured whether selection improves estimation and mostly found that it does not; line (2) measured whether accumulation achieves coverage and found that it does.

4.1 Line (1): thirteen pre-registered experiments

Each experiment in bio-3d-vision declared, before it ran, what would falsify its hypothesis and what would mean it had asked the wrong question. The results below are condensed from TR-26-09; experiment identifiers are those of the repository.

Question	Outcome	Experiment
Does the gaze objective matter?	No. Two objectives agreed on zero of 144 picks, a median 115 px apart, with indistinguishable results	exp001
Does posterior variance beat the validity mask?	Only marginally. Knowing *where* measurement is possible was worth 65× knowing *how well* on median error, 3.5× on the tail	exp002
Does verging to acquire help?	No. Median error rose 7.5×; re-verging halved the tail and added 21.7% coverage at 2.6× the hypotheses	exp003
Does the rendered scene match the analytic fixture?	Not at discontinuities: the fixture was 23× worse in the tail and 1.6× worse in the median	exp004
Were earlier results artefacts of the fixture?	The artefact carried the effects rather than masking them	exp005
Do policy results hold on rendered data?	Yes, and the variance channel recovered more strongly than on the fixture	exp007
How much occlusion makes the question visible?	Sharpest near a tenth of the surface occluded; indiscriminate above a sixth	exp008
What does violating epipolar rectification cost?	More than rectifying, at 9.5× the matcher time	exp009
Does closing the perception–action loop help?	Yes, at 41× the wall-clock cost, 91% of it rendering	exp010
Does that hold at more seeds and levels?	Direction yes, margin no; coverage exceeded accuracy by at least 33× and up to 341×	exp011
What did the inherited statistical bar test?	Not what it had been read as testing; affected verdicts were re-analysed	exp012
Does foveal weighting help on a complete sphere?	No: strictly lossy on every seed, leaving about four fifths of the sphere at the prior	exp013
Is the closed-loop gain allocation or scaling?	Re-acquisition of fresh images, away from discontinuities; open at them	exp014

Three further measurements shape the reading of that table:

- **The stimulus defect.** At depth discontinuities the warped random-dot fixture produced texture with no consistent disparity. 78.5% of those pixels passed the validity test and 21.7% of them were wrong by more than two pixels. Pixels within ten of a discontinuity were 38% of valid pixels and carried 82% of the squared error.
- **Policy lockup.** Without inhibition of return, the inherited maximum-variance loop spent 10 of 18 fixations on a single pixel where the matcher had hallucinated a disparity.
- **No acquisition on a complete sphere.** Rotating an eye about a fixed centre returns an identical capture from a complete spherical sensor, verified against Blender to 80 µm.

TR-26-09 summarises the pattern in one sentence: availability dominates selection. Its authors also question their own framing, concluding that the project may have posed a budget problem as an estimation problem.

4.2 Line (2): a reproducible 360° baseline and a foveated sensor

f3d-vision's official reproduction ran in Blender's Classroom with a fixed head, 64-spp renders and oracle correspondence. Its headline numbers are measured, not projected.

Quantity	Value
Fixations to the stopping rule	542 (322 local, 219 global saccades)
Sphere seen / with reconstructed depth	99.00% / 98.28% of 4π
First-hit reference solid angle within 12 mm	98.32% (98.52% of reference cells)
Distance to nearest surfel, covered cells	median 0.73 mm, 90th percentile 2.24 mm
Final map	16,517,811 surfels; about 1 GB, 2.2 GB per run
Wall time	733 s, of which rendering 413.5 s and fusion 257.8 s
Reproduction	23 of 23 strong-equivalence checks pass, bitwise
Truth firewall	3,813 files opened during control, none from the reference

Coverage grew in three phases: local walks alone to 56.06% by fixation 200; a mixed phase to 96.74% by fixation 450; then 92 global jumps into small holes at about 0.025% of the sphere each.

fov-3d-vision's Phase A measured the sensor side, monocularly:

- The foveated OSL camera renders at 50.8× fewer rays than uniform sampling of the same field.
- At equal rays, foveation beats uniform sampling on the fixated targets at every budget: 0.183 against 0.351 at the full profile's largest budget, and 0.146 against 0.365 from the first fixation.
- Over the whole sphere, uniform sampling wins at every budget below the largest, where the two are level. Fifty fixations cover 58% of the sphere, and the periphery between targets is charged for its blur.
- The warp parameter E_2 trades one objective against the other: at equal rays $E_2 = 4°$ wins the fixated targets and $E_2 = 1°$ the covered sphere. The choice is left to the objective (decision D11).

- A fixation costs 15 ms (small profile) or 140 ms (full profile) on an RTX 4090.

5 Strengths and weaknesses

Each line is strongest exactly where the other is weakest: line (1) on the estimator, line (2) on the sensing regime and the system around it.

Criterion	(1) Active Bayesian	(2) Foveal
Theoretical grounding	Strong: a six-layer framework with explicit equations and three closures	Light: a design principle and an explicit loop, motivated by Yarbus, Aloimonos, Ballard and Schwartz
Realism of measurement	Real matcher; confronts occlusion and confident-wrong matches	Oracle correspondence; matching errors removed by construction
Uncertainty	First-class: per-cell variance, propagated through the depth Jacobian	Absent: support counts only
Sensing regime	Unlimited uniform sampling; complete sphere in bio-3d, where a saccade acquires nothing	Narrow field, so every saccade acquires new directions
Foveation	Weighting on a uniform sensor, measured as strictly lossy	f3d: hard crop of the core; fov-3d: true variable-resolution sampling, 50.8× fewer rays
Gaze policy	Information-driven, reads the belief; locks up without inhibition of return	Coverage-driven, ignores the map; simple, deterministic, never stalled
Scale	Small analytic and rendered stimuli	A furnished 360° room, 16.5 million surfels
Outcome	Mostly negative, with portable findings	Positive proof of concept in a controlled setting
Rigour	Pre-registration, decision and measurement ledgers, forward-only history	Frozen baseline, truth firewall, bitwise reproduction
Efficiency	Closed loop costs 41×; heuristic, uncalibrated variance	Long hole-filling tail; about 1 GB map; periphery rendered and then discarded
Maturity of integration	Complete as a study; repository closed	The two repositories not yet joined; fov-3d is still monocular

5.1 Line (1)

The main strength of line (1) is that it asked the hard question honestly. It used a real matcher, pre-registered every test, and recorded corrections as amendments. That discipline turned negative results into transferable findings:

- **Masked-controller blindness.** A controller driven by a statistic over valid measurements cannot see the error that would correct it, because validity has already discarded the evidence. This holds wherever a validity mask sits upstream of a control loop.
- **Linearization as a confidence filter.** Expanding depth about the fixated disparity inflates the variance of estimates far from the fixated depth. That suppresses the confident-wrong matches that stereo confidence fails to report.
- **A synthetic stimulus is an instrument.** It must be characterised before it measures anything, because its worst behaviour tends to coincide with the regime of interest.

Its main weakness is the regime it tested. With unlimited uniform sampling, choosing where to attend can only re-weight information already bought. On a complete sphere, rotating the eyes changes nothing at all.

5.2 Line (2)

The main strength of line (2) is that it works end to end at scale, in the regime where looking matters, and that every component sits behind a narrow interface. The truth-stripped correspondence interface is the clearest case: a natural matcher can replace the oracle with nothing downstream changed. Its reproducibility makes it usable as a reference against which one change at a time can be measured.

Its main weakness is that the oracle removes precisely what line (1) found most consequential. Under perfect correspondence, the claim that a narrow sensor swept over a sphere eventually covers it is close to guaranteed, so the result validates the loop more than the vision. The map carries no uncertainty, the policy never reads it, and fov-3d-vision's foveated sensor has not yet been placed inside f3d-vision's loop.

6 Discussion: from estimation to availability

Read in order, line (2) is the corrective inheritance of line (1): it was built in the regime that line (1)'s results pointed to.

6.1 Two inversions

The architectures are mirror images on two axes.

- In line (1) the *policy reads the model* (the depth belief), but in bio-3d-vision the *sensor already sees everything*. Selection has nothing new to select.
- In line (2) the *sensor is narrow*, so each saccade acquires data, but the *policy ignores the model*. Gaze is driven by a coverage record and the current fixation's valid-geometry mask, never by the reconstructed map.

Each line therefore holds the half the other lacks. Line (1) has an informative model and no reason to move; line (2) has every reason to move and no model guiding where.

6.2 The negative results as a design brief

The conclusions of TR-26-09 read almost as a specification for line (2):

- "Availability dominates selection" corresponds to f3d-vision's coverage-driven policy and its stopping rule on the fraction of the sphere seen.
- The one positive result, closing the loop, was worth 33× to 341× more as coverage than as accuracy. f3d-vision is a system whose objective *is* coverage.
- "It may be we posed a budget problem as an estimation problem" corresponds to fov-3d-vision's premise that foveation is a bandwidth decision, and to its target of the same posterior for less computation.
- "Variable-resolution sampling is the priority" is fov-3d-vision's Phase A.

6.3 A tension inside line (2)

The two halves of line (2) currently optimise different objectives. fov-3d-vision found that foveation wins on fixated targets but loses to uniform sampling over the whole sphere at all but the largest budget. f3d-vision's objective is whole-sphere coverage. Placing the foveated camera inside the present explorer may therefore not pay. It should pay once the policy aims the fovea at specific places where precision matters, such as depth edges or poorly supported surfaces, while the periphery handles coverage.

6.4 What line (1) predicts for line (2)

Three findings of line (1) apply directly once line (2) replaces its oracle:

- **Masked-controller blindness** applies to f3d-vision's local saccade score, whose `edge_support` term is computed over valid geometry. With a natural matcher, surfaces that fail to match will look like the absence of surface.
- **The stimulus as instrument** applies to the oracle itself, which is an idealised instrument. Its idealisation is concentrated at occlusion boundaries, the same place the fixture's artefact lived.
- **Uncertainty is a refinement, not a driver.** Adding an information-gain term to f3d-vision's policy is worth testing, but line (1) predicts a second-order effect relative to availability.

6.5 Complementarity

The lines are best seen as two halves of one system. Line (1) supplies the estimator: a real matcher, calibrated uncertainty, the vergence–scale relation, and a catalogue of failure modes. Line (2) supplies the sensing regime and the system around it: a narrow, acquiring fovea; budgeted variable-resolution sampling; persistent head-frame memory; and a reproducible playground in which one component can be changed at a time. A companion report proposes an architecture that joins them.

7 Conclusion

The two lines are not rival answers to one question; they are answers to two questions that a complete active vision system must ask together.

Line (1) asked whether choosing where to look improves what a stereo estimator recovers. Under unlimited uniform sampling, and on a sensor that already sees every direction, the answer was largely no: what could be measured mattered far more than how cleverly measurements were chosen. Line (2) asked whether a narrow binocular fovea with persistent memory can reconstruct a whole scene. In a controlled synthetic setting with perfect correspondence, the answer was yes, reproducibly.

Neither result stands alone. Line (2)'s success depends on an oracle that hides the difficulties line (1) documented, and line (1)'s failure followed from a sensing regime that line (2) avoids. The next step is a system that puts line (1)'s estimator inside line (2)'s loop, with a variable-resolution sensor and a declared budget. The companion report, *Toward a Unified Architecture for Bio-Inspired Active 3D Vision*, sets out that architecture.

References

1. J. Aloimonos, I. Weiss and A. Bandyopadhyay. Active vision. *International Journal of Computer Vision* 1(4):333–356, 1988. doi:10.1007/BF00133571.
2. R. Bajcsy. Active perception. *Proceedings of the IEEE* 76(8), 1988. doi:10.1109/5.5968.
3. D. H. Ballard. Animate vision. *Artificial Intelligence* 48(1):57–86, 1991. doi:10.1016/0004-3702(91)90080-4.
4. B. Julesz. Binocular depth perception of computer-generated patterns. *Bell System Technical Journal* 39(5):1125–1162, 1960.
5. 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.
6. A. L. Yarbus. *Eye Movements and Vision*. Plenum Press, 1967.
7. L. Velho (with GPT). A framework for stereoscopic 3D vision: active Bayesian inference with nested oculomotor control. Working draft, VISGRAF/IMPA, August 2026.
8. L. Velho (with GPT). Bioeye: a bio-inspired active Bayesian stereo architecture. Technical report, VISGRAF/IMPA, September 2026.
9. L. Velho and Claude. Experiments on active 3D vision. Technical Report TR-26-09, VISGRAF/IMPA, 2026.
10. 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.

Repositories

- visgraf/bioeye
- visgraf/active-stereo (not publicly reachable on 17 October 2026)
- visgraf/bio-3d-vision
- visgraf/fov-3d-vision
- visgraf/f3d-vision, tag `greedy-foveal-playground-baseline`
- visgraf/w3d-scenes (archived)

Appendix A: Provenance

This report was drafted by Claude (Anthropic, Opus 5.5) at the request of L. Velho, who directed its scope. It is a synthesis: it introduces no new measurements. Every number is quoted from the technical report or repository README that produced it, read on 10 and 17 October 2026. Where sources disagree on a detail, such as the number of experiments in active-stereo, the detail is omitted rather than chosen.