Bioeye

A Bio-Inspired Active Bayesian Stereo Architecture

Fixation-centered geometry • uncertainty-aware fusion • active gaze

Luiz Velho

IMPA

GPT

Open.AI (collaborator)

Abstract

Bioeye recasts stereo depth estimation as a closed-loop sensing problem rather than a single dense reconstruction pass. At each iteration, the system selects a fixation from the current depth uncertainty, establishes a vergence-centered reference depth, measures residual disparity with foveated emphasis, converts that residual into metric depth, and fuses the observation with an accumulated Bayesian belief. The key local relation,

$$Z \approx D - \frac{D^2}{I}\eta,$$

is the first-order expansion of pinhole stereo depth around the fixation distance D. It makes vergence part of the metric calibration and exposes how depth uncertainty grows rapidly with distance. This report consolidates the method, the seven operations discussed in the project conversation, their mapping to layers L0–L5, the two principal feedback loops, and the current Python prototype's implementation boundaries. It also separates implemented behavior from conceptual extensions and proposes a validation agenda.

Keywords: active vision; binocular stereo; vergence; foveation; Bayesian fusion; depth uncertainty; gaze control.

> **Core idea**
>
> Bioeye spends sensing effort where the current depth belief is least certain, while using fixation and vergence to keep local disparity-to-depth conversion well conditioned.

1. Scope and evidence base

This report synthesizes the earlier Active Vision Duo discussion of the VISGRAF Bioeye repository and cross-checks it against the public repository inventory available in September 2026. The source repository is described by VISGRAF as a Python project for biological 3D vision [VISGRAF Laboratory, 2026a]. A measured migration inventory records the Bioeye revision `e90817018e629ce1cd23af0c9e8bc4a6aa15daff`, 20 tracked files, four Python files, no automated tests, and a 432-line demonstration that contains all six architectural layers [VISGRAF Laboratory, 2026b].

Two levels of description are kept distinct. The implementation level covers the rectified, parallel-pinhole demo, its disparity front ends, scalar fixation disparity, foveated measurement, precision-weighted fusion, and output plots. The architectural level generalizes these operations into retinal encoding, uncertainty-bearing evidence, oculomotor control, Bayesian state, and active gaze. The latter is a faithful interpretation of the loop, not a claim that every biological mechanism is already simulated.

2. System overview

A conventional stereo pipeline typically estimates disparity once and immediately converts it to depth. Bioeye instead maintains a spatial belief over depth. The belief determines where to look next; the selected fixation determines vergence and the local metric scale; new stereo evidence updates the belief; and the revised uncertainty starts the next cycle. The architecture is therefore partly feed-forward and essentially recurrent.

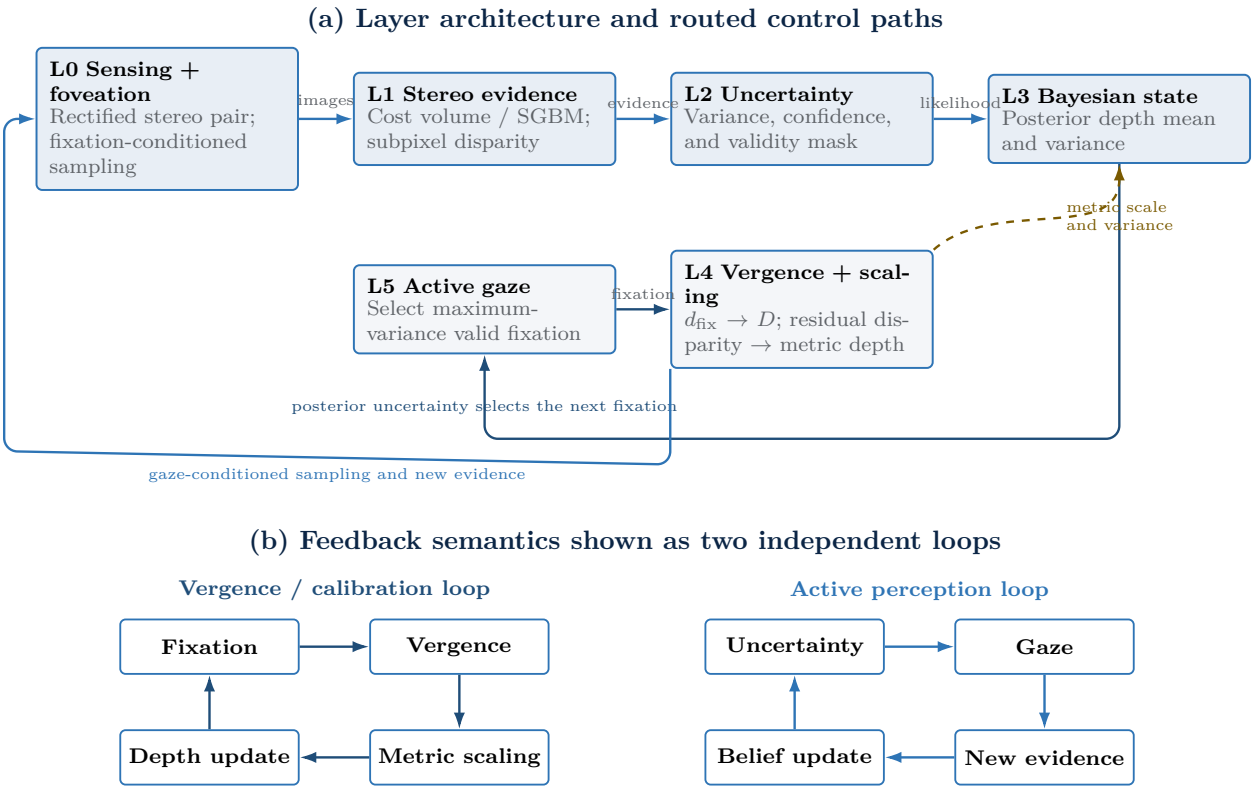


Figure 1: *Improved Bioeye architecture and feedback interpretation. Panel (a) separates the feed-forward evidence path from the routed control paths. Panel (b) expands the two loops independently so their semantics can be read without crossing arrows.*

3. Seven operations and their layer mapping

The seven operations below are ordered as a computational cycle. Some operations cross a layer boundary because fixation changes both sensory sampling and the disparity-to-depth transform. This is why Figure 1 should not be read as a simple neural-network stack.

Table 1: *Seven Bioeye operations and their primary layer paths.*

Op.	Function	Operational meaning	Layer path
1	Acquire and calibrate	Read a rectified left/right pair and the pinhole parameters: focal length f and interocular baseline I.	L0
2	Estimate disparity evidence	Build stereo costs, decode subpixel disparity, and retain correspondence evidence. The demo supports block SSD and SGBM front ends.	L1
3	Estimate reliability	Derive per-pixel variance or confidence and validity, including left-right consistency where available.	L2
4	Select fixation	Choose the highest-variance valid location, $q_t = \arg\max_x \sigma_t^2(x)$, to reduce expected uncertainty greedily.	L5 $\leftarrow$ L3
5	Set vergence and foveate	Read the fixation disparity d_{fix}, infer $D = fI/d_{\text{fix}}$, and emphasize a neighborhood around q_t.	L4 $\rightarrow$ L0
6	Recover metric depth	Convert residual disparity about fixation to depth and propagate measurement variance through the local Jacobian.	L4 $\rightarrow$ L3
7	Fuse and repeat	Combine the measurement with the current Gaussian depth belief by precision weighting; update uncertainty and choose the next fixation.	L3 $\rightarrow$ L5

4. Layer responsibilities

The layer split separates quantities that have different semantics and failure modes. In particular, L2 owns uncertainty as a first-class signal; L3 owns accumulated world state; L4 owns the fixation-dependent metric closure; and L5 owns action selection.

Table 2: *Responsibilities and outputs of layers L0–L5.*

Layer	Responsibility	Contents	Principal output
L0	Sensing and foveation	Rectified binocular images, calibration, fixation-conditioned sampling or weighting.	Stereo pair and foveated support
L1	Binocular evidence	Disparity hypotheses and matching costs from block SSD, SGBM, or a generalized retinal matcher.	Disparity mean or cost volume
L2	Uncertainty and validity	Confidence, variance, left-right consistency, occlusion or invalidity masks.	Measurement variance and valid mask
L3	Bayesian depth state	Per-pixel posterior mean and variance, accumulated across fixations; evaluation metrics such as RMSE.	$\mu_t(x), \sigma_t^2(x)$
L4	Vergence and metric scaling	Scalar fixation disparity in the prototype; fixation distance, local Jacobian, metric depth, and variance.	D_t, J_t, z_t, τ_t^2
L5	Active gaze policy	Select the next fixation from posterior uncertainty; the current policy is greedy and myopic.	q_{t+1}

> **Important**
>
> L4 feeds back into both L0 and L3: fixation changes what is sampled and also changes how residual disparity is interpreted metrically. L5 closes the larger perception-action loop.

5. Fixation-centered stereo geometry

5.1 Conventional depth-disparity relation

For a rectified parallel stereo rig, let f be focal length in pixels, I the interocular baseline in metric units, d the horizontal disparity in pixels, and Z the corresponding depth in the same metric units as I. The pinhole relation is

$$d(Z) = \frac{fI}{Z} \quad \Longleftrightarrow \quad Z(d) = \frac{fI}{d}. \tag{1}$$

At the current fixation distance D, the reference disparity is $d_{\text{fix}} = fI/D$. A scene point generates the residual pixel disparity $\Delta d = d - d_{\text{fix}}$. Positive residual disparity denotes a point nearer than the fixation plane under the usual left-minus-right sign convention.

5.2 Taylor expansion around fixation

Differentiate $Z(d) = fI/d$ and evaluate the derivative at d_{fix}:

$$\left.\frac{\mathrm{d}Z}{\mathrm{d}d}\right|_{d_{\text{fix}}} = -\frac{fI}{d_{\text{fix}}^2} = -\frac{D^2}{fI}. \tag{2}$$

The first-order approximation for pixel residual disparity is therefore

$$Z \approx D - \frac{D^2}{fI}\Delta d. \tag{3}$$

Define normalized residual disparity $\eta = \Delta d/f$. It is dimensionless and, for small angles, can be interpreted as an angular disparity in radians. Substitution gives the vergence-centered form discussed in the project conversation:

$$\boxed{Z \approx D - \frac{D^2}{I}\eta.} \tag{4}$$

The same result follows from the exact normalized expression

$$Z = \frac{DI}{I + D\eta}. \tag{5}$$

Expanding $1/(1 + D\eta/I)$ around $\eta = 0$ yields Equation (4), with leading truncation error of order $D^3\eta^2/I^2$. The approximation is locally reliable when $|D\eta/I| \ll 1$. Foveal confinement is therefore not merely a computational convenience: it helps keep the linearization close to its expansion point.

5.3 Units, sign, and conditioning

Table 3: *Notation used in the fixation-centered derivation.*

Symbol	Units	Meaning
f	pixels	Focal length of the rectified pinhole camera
I	metres, or any metric unit	Interocular or stereo baseline
D, Z	same units as I	Fixation depth and scene depth
$d, d_{\text{fix}}, \Delta d$	pixels	Absolute, fixation, and residual disparities
$\eta = \Delta d / f$	dimensionless	Normalized residual disparity; approximately angular for small angles

Equation (4) has the expected sign: $\eta > 0$ implies $Z < D$, while $\eta < 0$ implies $Z > D$. It also shows the conditioning problem of stereo at distance: the same normalized disparity error causes a metric error proportional to D^2.

6. Uncertainty propagation and Bayesian accumulation

6.1 Metric uncertainty

Let η be a noisy measurement with variance σ_η^2. Linear propagation through Equation (4) uses the Jacobian $J = \partial Z / \partial \eta = -D^2 / I$:

$$\sigma_Z^2 \approx J^2 \sigma_\eta^2 = \frac{D^4}{I^2} \sigma_\eta^2. \tag{6}$$

In pixel units, $\sigma_Z^2 \approx [D^4/(f^2 I^2)]\sigma_{\Delta d}^2$. Thus depth variance grows as D^4 when disparity noise is approximately constant, a behavior explicitly pinned in the successor active-stereo tests [VISGRAF Laboratory, 2026b].

6.2 Precision-weighted fusion

For each valid pixel x, represent the accumulated depth belief at iteration t as a Gaussian $\mathcal{N}(\mu_t, \sigma_t^2)$ and the new fixation-conditioned measurement as $\mathcal{N}(z_t, \tau_t^2)$. Independent Gaussian fusion is

$$\sigma_{t+1}^{-2} = \sigma_t^{-2} + \tau_t^{-2}, \tag{7}$$

$$\mu_{t+1} = \sigma_{t+1}^2 \left(\frac{\mu_t}{\sigma_t^2} + \frac{z_t}{\tau_t^2} \right). \tag{8}$$

This update gives greater weight to the more precise source and monotonically reduces posterior variance when the measurement is valid and its uncertainty is meaningful. Invalid or inconsistent stereo matches should contribute zero precision rather than a fabricated depth. The practical quality of the posterior therefore depends as much on uncertainty calibration and masking as on the disparity mean.

6.3 Generalized retinal representation

The companion active-vision discussion expressed local stereo evidence through Gaussian retinal tokens. For a receptive-field kernel ϕ_i and images L^L and L^R, define

$$r_i^L = \int L^L(x)\,\phi_i(x)\,\mathrm{d}x, \tag{9}$$

$$r_i^R(Z) = \int L^R(x+\delta(Z))\,\phi_i(x)\,\mathrm{d}x, \qquad \delta(Z) = \frac{fI}{Z}. \tag{10}$$

A depth hypothesis can be selected by minimizing

$$E_i(Z) = \left\| r_i^L - r_i^R(Z) \right\|^2, \tag{11}$$

or equivalently by maximizing $C_i(Z) = \langle r_i^L, r_i^R(Z) \rangle$. This formalism clarifies how foveated, variable-resolution measurements could replace the current block/SGBM front end without changing the Bayesian and control layers. It should be treated as an architectural extension; the measured Bioeye demo uses conventional dense stereo front ends.

7. The two feedback loops

7.1 Vergence-to-metric-scaling loop

Fixation selects d_{fix}; d_{fix} determines $D = fI/d_{\text{fix}}$; D determines both the local depth Jacobian and the conversion of residual disparity to metric depth; the new metric estimate alters the posterior that will govern later fixations. This is the inner calibration loop. Vergence is therefore not only a motor command: it is part of the measurement model.

7.2 Uncertainty-driven active-gaze loop

Posterior uncertainty identifies where the model lacks information. The current policy chooses the valid pixel with maximum variance. A new fixation changes the foveated support and vergence reference, producing new evidence. Bayesian fusion reduces uncertainty locally, shifting attention toward another region. This is the outer perception-action loop:

$$\text{uncertainty} \rightarrow \text{gaze} \rightarrow \text{new evidence} \rightarrow \text{belief update}.$$

> **Interpretation**
>
> The two loops are nested. L5 decides where to acquire evidence; L4 decides how that fixation calibrates the evidence; L3 records the consequence and exposes the remaining uncertainty.

8. Reference algorithm

A single Bioeye run can be summarized as the following control procedure:

1. Compute dense disparity, measurement uncertainty, and a valid-data mask from the stereo pair.
2. Initialize the depth posterior $\mu_0(x), \sigma_0^2(x)$, assigning high variance where depth is unknown.

3. Select $q_t = \arg\max_x \sigma_t^2(x)$ over valid candidate fixations.
4. Set $d_{\text{fix},t}$ from the disparity at q_t and compute $D_t = fI/d_{\text{fix},t}$.
5. Form a foveated measurement around q_t and compute residual disparity relative to $d_{\text{fix},t}$.
6. Convert residual disparity and variance to metric depth using the fixation-centered Jacobian.
7. Fuse the valid measurement with the posterior by precision weighting.
8. Record RMSE or another evaluation measure, then repeat until the fixation budget or stopping criterion is reached.

9. Correspondence with the Python prototype

The measured repository inventory identifies the following implementation anchors [VISGRAF Laboratory, 2026b]. Line numbers refer to the audited Bioeye revision and may move in later versions.

Table 4: *Implementation anchors in* `active_stereo_demo.py`.

Implementation anchor	Role in the method
`make_synthetic_scene` (47–77)	Creates fronto-parallel cards on a background and inverse-warps the left image to obtain a right image whose ground-truth disparity is $d_{\text{gt}} = fI/Z$.
`cost_volume`, `decode_disparity`, `lr_consistency` (112–160)	Box-aggregated SSD costs, winner-take-all decoding with parabolic subpixel refinement, and left-right consistency.
`front_end_block`, `front_end_sgbm` (168–202)	Alternative dense stereo front ends. SGBM variance is documented as a roughness heuristic rather than calibrated probability.
`scale_to_depth` (208–218)	Implements the vergence-centered scaling relation and Jacobian, first-order exact at fixation.
`ActiveStereo.step`, `run` (225–313)	Maximum-variance fixation, scalar vergence, foveated metric measurement, precision-weighted fusion, and RMSE history.
`vergence_scaling_experiment` (319–352)	Plots the claim that, under the experiment's error convention, fractional vergence error e produces relative depth error $(1+e)^2 - 1 \approx 2e$.
`save_result_fig` (370–395)	Produces four panels: posterior mean with scanpath, posterior standard deviation, absolute error, and RMSE versus fixation number.

10. Assumptions and limitations

- **Rectified parallel geometry.** The prototype uses $d = fI/Z$ throughout; it has no camera rotation, SO(3) representation, or ray-based geometry.
- **Scalar vergence state.** `ActiveStereo.vergence` returns a fixation disparity in pixels, not a three-dimensional gaze direction or binocular eye pose.
- **Local linearization.** Equation (4) is first-order exact at fixation and degrades with peripheral residual disparity; foveal confinement must be tested rather than assumed.
- **Uncertainty calibration.** A heuristic variance can guide attention but does not automatically justify Bayesian probabilities or credible intervals.

- **Greedy gaze.** Maximum posterior variance is myopic; it ignores motion cost, future observations, task relevance, occlusion, and multi-step information gain.
- **Limited biological scope.** The current design deliberately omits process dynamics, trans-saccadic remapping, accommodative-convergence coupling (AC/A), vertical-disparity cues, and a non-myopic policy [VISGRAF Laboratory, 2026b].
- **Prototype maturity.** The audited repository has no automated tests. The behavioral expectation that RMSE decreases with fixation count is falsifiable, but is not currently enforced by a Bioeye test suite because no such suite exists.

11. Validation agenda

The most useful next step is to convert the architectural claims into quantitative closures and regression tests.

Table 5: *Proposed validation targets.*

Validation target	Concrete check
Geometry closure	Round-trip $Z \to d \to Z$; verify Equation (4) at fixation and bound error versus eccentricity or residual disparity.
Variance law	Inject controlled disparity noise and confirm $\sigma_Z^2 \propto D^4$, including pixel and normalized-disparity conventions.
Uncertainty calibration	Measure negative log likelihood, empirical coverage, and calibration curves; reject invalid matches instead of assigning overconfident depth.
Active policy benefit	Compare maximum variance against random, raster, saliency-only, and oracle policies under a fixed fixation budget; report RMSE-versus-fixation AUC.
Ablation	Disable foveation, vergence scaling, or Bayesian accumulation separately to identify which component produces the gain.
Robustness	Test textureless regions, occlusions, repeated patterns, discontinuities, vergence perturbation, and real calibrated stereo images.
Geometry generalization	Replace scalar disparity fixation with rays and explicit eye/camera rotations while preserving the three closures: geometry, scaling, and uncertainty.

12. Conclusion

Bioeye's central contribution is architectural: it links binocular evidence, metric depth, uncertainty, fixation, and gaze in a single recurrent system. The vergence-centered approximation explains how local residual disparity becomes metric depth and why uncertainty grows strongly with distance. Precision-weighted accumulation makes multiple fixations useful, while posterior variance supplies a simple active policy. The current prototype demonstrates this logic in a deliberately constrained rectified-pinhole setting. Its next research milestone is calibrated uncertainty, falsifiable loop closures, and a geometry that can represent true binocular eye pose without losing the clean separation between sensing, inference, scaling, belief, and action.

References

VISGRAF Laboratory. Bioeye: Biological 3d vision. GitHub repository, 2026a. URL https://github.com/visgraf/bioeye. Accessed 2 September 2026.

VISGRAF Laboratory. Migration inventory—active-stereo and bioeye. Repository audit, 2026b. URL https://github.com/visgraf/bio-3d-vision/blob/main/docs/migration-inventory.md. Accessed 2 September 2026.