Nominal Vergence from Visual Data — architecture and tasks

Nov 6, 2026 · @Luiz Velho (with Claude Chat)

Goal

Build a routine that, given a proposed fixation direction ω, finds the vergence that fixates it from the stereo images alone, with no ground-truth depth. It returns the inverse depth $\hat{\rho}$ of the fixated point, its uncertainty σ_ρ, and a flag saying whether the fixation counts as measured.

The head is fixed and the eyes only rotate, at a baseline b apart. Write a direction in epipolar coordinates (θ, ϕ): θ picks a plane through the baseline, and ϕ is the angle from the baseline axis inside that plane. Both eyes share θ, so vergence is a rotation inside one plane. For a point on ω at inverse depth ρ, the two eyes need the in-plane angles

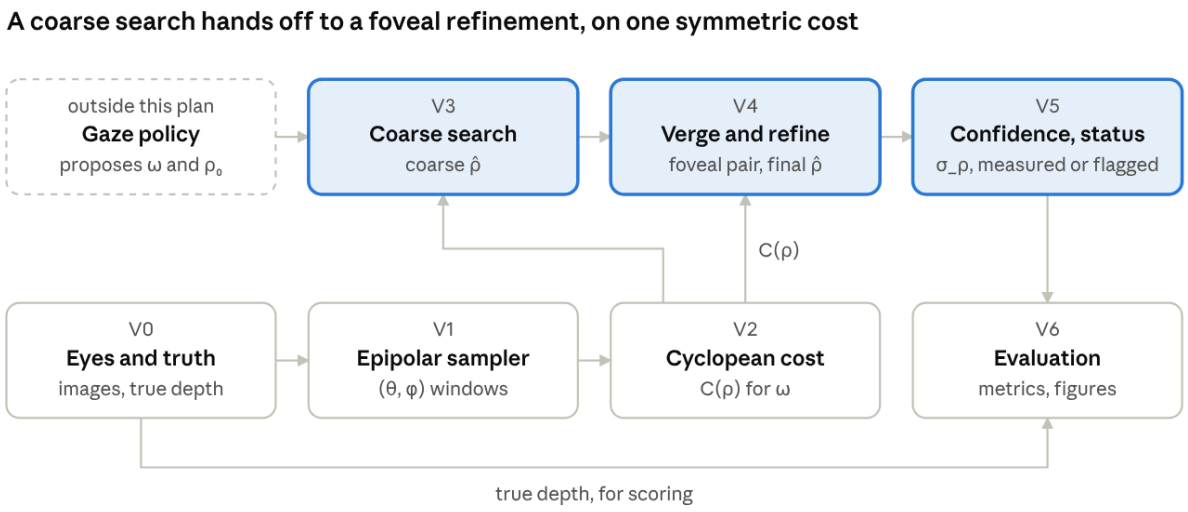
$$\varphi_L(\rho) = \operatorname{atan2}\big(\sin\varphi,\ \cos\varphi + b\rho/2\big), \qquad \varphi_R(\rho) = \operatorname{atan2}\big(\sin\varphi,\ \cos\varphi - b\rho/2\big)$$

and the vergence angle is $\gamma(\rho) = \phi_R(\rho) - \phi_L(\rho)$, about $b\rho \sin\phi$. Finding the vergence is therefore a one-dimensional search over ρ. The search must be symmetric: both eyes' windows move together, so the matched point stays on the cyclopean ray ω.

These definitions are the specification. How to search, how to refine and how to estimate confidence are left to the team.

Architecture

The routine runs once per proposed fixation: a coarse search, then a verge and a foveal refinement, then a confidence check. The top row is the procedure; the bottom row is the machinery it calls. The three highlighted modules are the core of the work.



architecture · 7 modules

V3 and V4 both call the same cost V2, at different levels: coarse for the search, foveal for the refinement. Because the eyes only rotate, any candidate vergence can be tried by resampling, without moving the eyes.

The modules exchange five data objects, which the team should fix in week one:

- **Fixation request:** $\omega = (\theta, \phi)$ and a prior ρ_o with its spread.
- **Eye pose:** the in-plane angle of each eye and the shared θ.
- **Window:** samples on an (θ, ϕ) grid, with eye and level.
- **Fixation result:** final $\hat{\rho}$, σ_ρ, status, and the verged foveal pair.
- **Metrics table:** one row per fixation.

What needs to be implemented

Seven modules, each defined only by what goes in, what comes out, and what it must do. Methods, parameters and thresholds are the team's choice.

Module	Input	Output	Must
V0 · Eyes and truth	Scene, baseline b, an eye direction for each eye, a resolution level	Left and right images for those directions; true depth along any ω	Render or resample at a coarse (peripheral) and a fine (foveal) level; rotate the eyes about their nodal points only
V1 · Epipolar sampler	An image, the eye, a direction (θ, ϕ), a level	A window sampled on an (θ, ϕ) grid	Align the window's rows with the epipolar direction, so both eyes share θ
V2 · Cyclopean cost	Fixation ω, a candidate ρ, a level	Matching cost $C(\rho)$	Compare the left window at $\phi_L(\rho)$ with the right window at $\phi_R(\rho)$; never anchor one eye and search the other
V3 · Prior and coarse search	ω, a prior ρ_o (from the belief, the coarse images, or a default)	Coarse estimate $\hat{\rho}$ and its uncertainty	Search a range of ρ around ρ_o on coarse data; land within the fine level's pull-in range
V4 · Verge and refine	ω, coarse $\hat{\rho}$	Verged foveal pair and final $\hat{\rho}$	Point the eyes at $\phi_L(\hat{\rho})$, $\phi_R(\hat{\rho})$; correct ρ on the foveal data until the residual central disparity is at noise level
V5 · Confidence and status	Costs and windows from V3 and V4	σ_ρ and a status: clean, repeated texture, depth edge, half-occlusion, near baseline axis, low texture	Turn angular precision into σ_ρ; mark a fixation as measured only when it is clean and σ_ρ is small enough

Module	Input	Output	Must
V6 · Evaluation	Results of many fixations, truth from V0	Metrics table and figures	Measure depth error, success rate against prior error, calibration of σ_ρ, and status against true failure cases

A scale to keep in mind for V3 and V4: at b = 63 mm, an inverse-depth error of 1 m^{-1} is about 3.6° of disparity, larger than a 2° fovea. That is why V3 has to come close before V4 can take over.

Deliverables

The work is judged by three comparisons over a set of proposed fixations in at least one scene. The team sets its own success criteria, and writes them down before each run.

1. **Accuracy.** The final $\hat{\rho}$ against the true depth along ω, for clean fixations, reported against distance and against the angle ϕ from the baseline axis.
2. **Pull-in.** Success rate as the prior error grows: how far from the truth V3 can start and still hand V4 a value it can finish.
3. **Honesty.** Whether σ_ρ predicts the actual error, and whether the status flags catch the true repeated-texture, edge and half-occlusion cases.

The team delivers the code for V0 to V6, the metrics table and figures for the three comparisons, and a short report of what worked and what did not.

Using it with the window plan. The verged foveal pair from V4 is the pair record that the window plan's M0 needs, with ω and its vergence. Run together, this routine supplies the fixations, and the window plan reads them. Run apart, each stands alone: the window plan can verge from true depth instead, and this plan ends at a verged pair with its σ_ρ and status.