Finding the Nominal Vergence from Visual Data in Active Foveal Vision

Luiz Velho (IMPA), with Claude Chat (Anthropic)

In the shared epipolar plane, the vergence that fixates a point is exactly that point's binocular disparity. So “find the nominal vergence for a fixation ω” is a one-dimensional correspondence search along the epipolar circle that contains ω, and you finish with a servo step that drives the foveal disparity to zero.

Geometry: everything happens in one plane

Take the head (cyclopean) origin C and eyes at L, R = C ∓ (b/2)$\hat{x}$, with $\hat{x}$ the baseline axis. Write directions in epipolar coordinates (θ, φ). θ picks a plane through the baseline, and φ is the angle from the baseline axis within that plane. A proposed fixation is a cyclopean direction ω = (θ, φ), and the fixation point P = C + zω is unknown only through its depth.

L, R, C and every candidate P lie in the plane θ. With inverse depth ρ = 1/z, the in-plane gaze angles each eye needs are:

$$\phi_L(\varrho) = \operatorname{atan2}(\sin\phi,\ \cos\phi + b\varrho/2), \qquad \phi_R(\varrho) = \operatorname{atan2}(\sin\phi,\ \cos\phi - b\varrho/2)$$

Both eyes share the same θ. That means there is no vertical disparity, and vergence is a pure in-plane rotation. This is why Helmholtz-style eye coordinates are the natural ones here.

By the exterior-angle theorem in triangle L R P, the angle at R equals the angle at L plus the angle at P:

$$\gamma(\varrho) = \phi_R(\varrho) - \phi_L(\varrho) \quad \text{(exact)}, \qquad \gamma \approx b\,\varrho\,\sin\phi$$

Here γ is the vergence angle, the angle the baseline subtends at P. The left side, $\varphi_R - \varphi_L$, is also P's angular disparity measured in head-fixed epipolar coordinates. So the vergence needed to fixate P equals P's epipolar disparity, and the set of candidate vergences for ω is a one-parameter family indexed by ρ.

Measuring it from the images

A rotation about the nodal point is pure reprojection with no parallax. With the head fixed, you can therefore evaluate any candidate vergence on data you already have, without moving the eyes. Define a symmetric, cyclopean matching cost:

$$C(\varrho) = 1 - \mathrm{NCC}(W_L(\theta, \phi_L(\varrho)),\ W_R(\theta, \phi_R(\varrho)))$$

The windows are expressed in (θ, φ), so their rows line up along the epipolar direction. The nominal vergence is $\gamma(\hat{\rho})$ with $\hat{\rho} = \operatorname{argmin} C(\rho)$.

The search has to be symmetric. If you instead anchor a window at the left eye's direction and search the right eye, you get the disparity of whatever surface the left eye happens to hit, not of the point on ω. Moving both windows together along their epipolar circles keeps the triangulated point on the cyclopean ray.

The active procedure

1. **Prior.** Start from ρ_0: the belief at ω if that region has been seen, otherwise the coarse peripheral levels of the current pair if ω falls inside them, otherwise a default.
2. **Coarse search.** Minimize $C(\rho)$ over a range around ρ_0 on the coarse data and get $\hat{\rho}$. Use the curvature at the minimum to get its uncertainty.
3. **Verge and sample the foveae** at $\varphi_L(\hat{\rho})$, $\varphi_R(\hat{\rho})$, with the shared θ.
4. **Fine correction.** Re-evaluate $C(\rho)$ on the foveal data, in a narrow interval around $\hat{\rho}$. If the minimum is at the current ρ, the residual central disparity is zero and you are done. Otherwise update ρ and repeat. Because γ is nearly linear in ρ, one correction is usually enough.

The real constraint is pull-in range. The coarse error must land inside the fovea's useful search width. At $b = 63$ mm, an error of $\Delta\rho = 1\ \mathrm{m}^{-1}$ is already about 3.6° of disparity, larger than a 2° fovea. So step 2 needs to get within roughly $\Delta\rho \lesssim 0.5\ \mathrm{m}^{-1}$ (about 1.8°). This is the vergence version of coarse-to-fine.

The classical alternative to the explicit search is a disparity servo on the fovea, such as phase difference of Gabor filters or Lucas–Kanade along φ. It returns the residual δ directly and lets you update $\gamma \leftarrow \gamma + \delta$. This is cheaper per step but has a smaller pull-in range, and it has the asymmetry problem above unless you apply it to symmetric windows.

Confidence and failure modes

- **Precision.** The disparity variance is roughly $\sigma_\gamma^2 \approx \sigma_n^2 / \Sigma(\partial I/\partial\varphi)^2$, summed over both windows, and $\sigma_\rho = \sigma_\gamma / (b \sin\varphi)$. Only texture with a component along the epipolar direction helps, so edges parallel to the epipolar circle are blind. As ω approaches the baseline axis ($\varphi \to 0$), sensitivity vanishes.
- **Repeated texture** produces several minima. The prior and the coarse levels have to break the tie.
- **ω on a depth edge** gives two minima, one per surface. Prefer the one supported at the window center, or nudge ω slightly with a micro-saccade.
- **Half-occlusion**, where the point on ω is visible to only one eye, has no valid vergence. You can detect it from a weak peak or a left–right inconsistency. Keep the prior, flag the fixation, and let the gaze policy treat it as unresolved.

In the pipeline, this is what replaces the ground-truth depth $\hat{z}$ in the pair specification $(\omega, \hat{z})$. The policy proposes ω, the vergence routine returns $\hat{\rho}$ with σ_ρ from the images alone, and the gate above decides whether that fixation counts as measured.