A Model of Binocular Gaze Estimation

Miles Hansard & Radu Horaud

INRIA Rhône-Alpes, 655 Avenue de l'Europe, Montbonnot 38334, France.

miles.hansard@inrialpes.fr radu.horaud@inrialpes.fr

1. Problem Statement

Estimation How can the 3-D position of the binocular fixation-point be estimated from local disparity information?

Representation How should the binocular fixation-point be represented in the visual system?

2. Assumptions

Calibration Angles between visual rays can be computed from retinal separations.

Fixation The optic axes meet at a point in space.

Perspective We use the standard 'pinhole camera' imaging model.

3. Epipolar Geometry

- Consider a pair of matched points $\boldsymbol{q}_\ell$ and $\boldsymbol{q}_r$.
- The 3-D scene point must lie on a ray starting at the left optical centre $\boldsymbol{c}_\ell$, and passing through $\boldsymbol{q}_\ell$.
- If this ray is projected into the right image, then $\boldsymbol{q}_r$ must lie on the resulting line.
- A similar relation holds for projection of the ray through $\boldsymbol{q}_r$, and over all matched points.

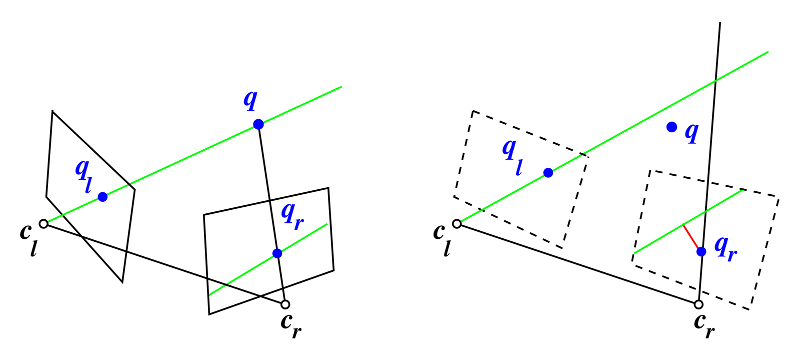


Figure 1: *Left: The ray through $\boldsymbol{q}_\ell$ projects to an epipolar line in I_r. The matching point $\boldsymbol{q}_r$ is on the line. Right: Orientations of the eyes are incorrectly estimated. The point $\boldsymbol{q}_r$ does not lie on the epipolar line; the (minimum) error is shown in red.*

- If the orientations of the eyes are not correctly estimated, then in general, the epipolar geometry will be inconsistent with the configuration of matched points.

4. Feature Extraction & Matching

- Apply derivative-of-Gaussian filters ∇_σ.
- Represent local image structure by 2×2 matrix $S(x,y) = \nabla_\sigma I \, (\nabla_\sigma I)^\top$.
- $S(x,y)$ is smoothed at scale τ. The response $C(x,y) = \det S_\tau \big/ \operatorname{tr} S_\tau$ has maxima at distinctive image-points (q_x, q_y).
- Consider the colour of I_ℓ around $\boldsymbol{q}_\ell^i$, vs. the colour of I_r around $\boldsymbol{q}_r^j$.
- Difference between the two colours is put into a matching-table, f_{ij}.
- Compute the minima m_ℓ^i in each row i, and m_r^j in each column j.
- Enforce uniqueness and compatibility, hence $\boldsymbol{q}_\ell^i$ matches $\boldsymbol{q}_r^j$ if $i = m_r^j$, $j = m_\ell^i$, and $f_{ij} < f_{\max}$.

5. Examples

- We use the freely available 'Tsukuba' stereo pair.

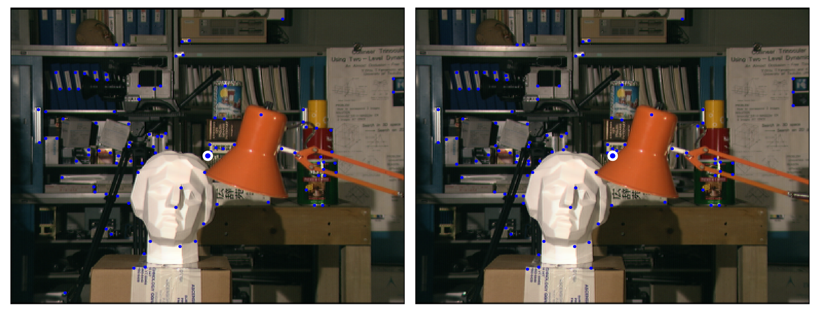


Figure 2: *Stereo image-pair used in the experiments. Example matched points are shown in blue. The fixation point (white) is at infinity.*

- In principle we need a data set which includes the left & right images obtained by fixating each scene-point that we test.
- In practice we need only one pair of images. The rotated views can be exactly synthesized by warping the source images.

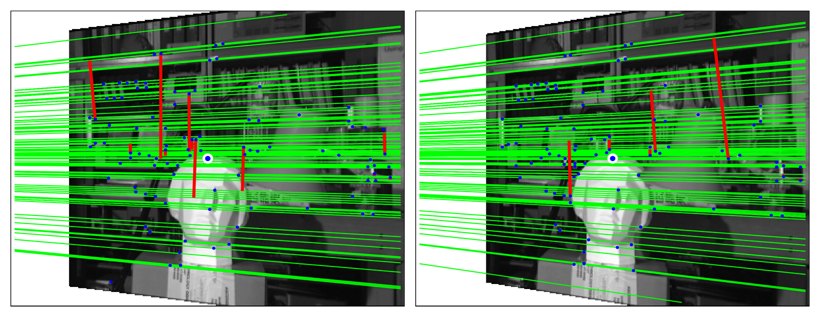


Figure 3: *Top: example of a 'fixating' image pair, with true epipolar lines shown in green. Red lines are nonzero point-line distances, caused by false feature matches.*

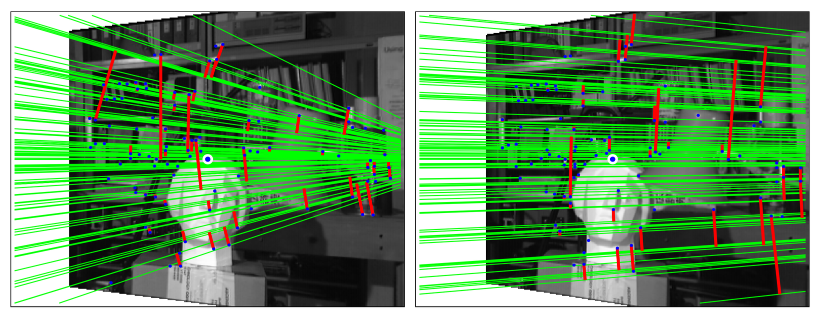


Figure 4: *The same images with an incorrect epipolar geometry. The number of nonzero point-line distances has increased (more red lines).*

6. Parameterization

- Orientation of the eyes is determined by the 3-D position of the fixation point $\boldsymbol{p}$.
- The fixation point is represented in Cyclopean spherical coordinates (α, β, ρ).

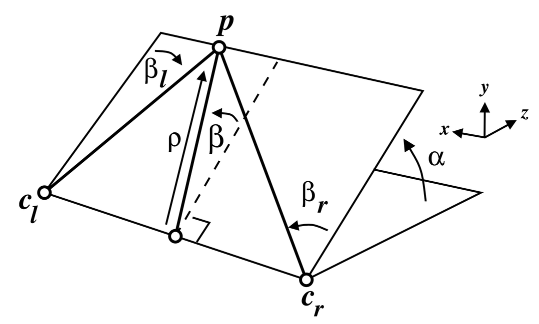


Figure 5: *A point $\boldsymbol{p}$ is fixated by eyes with optical centres $\boldsymbol{c}_\ell$ and $\boldsymbol{c}_r$. The whole arrangement is determined by the Cyclopean elevation α, azimuth β, and range ρ.*

- Epipolar geometry is a function of the relative orientation of the eyes; hence it is determined by Cyclopean azimuth β and range ρ.
- Note that the elevation angle α cannot be recovered from this parameterization.

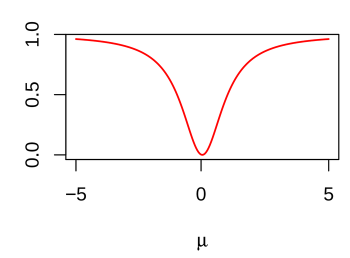
7. Cost Function

- If a feature-point $\boldsymbol{q}^k$ lies close to the corresponding epipolar line, then it is consistent with the gaze parameters.
- Test this constraint by finding the minimum distance μ of the point to the line. The line is represented by a point $\boldsymbol{q}'$ and a direction $\boldsymbol{d}$, hence

$$\mu^2 = \min_\lambda \left| \boldsymbol{q} - (\boldsymbol{q}' + \lambda \boldsymbol{d}) \right|^2.$$

- Wrongly-matched points are uninformative, and typically associated with large μ. Hence the cost function must be robust.
- We use the Geman-McClure function; for the k-th point, the penalty is:

$$E^k = \frac{\mu^2}{1 + \mu^2}.$$



- The final cost function is a sum over all n matched points, across both images:

$$E = \frac{1}{2n} \sum_k^n \left(E_\ell^k + E_r^k \right).$$

8. Gaze Localization Results

- Here we show examples of the epipolar cost function E, plotted over the horizontal plane $(x, 0, z)$.
- The plane is measured in human interocular-units; this means that $z_{50} \sim 50 \times 63\text{mm} = 3.15\text{m}$.

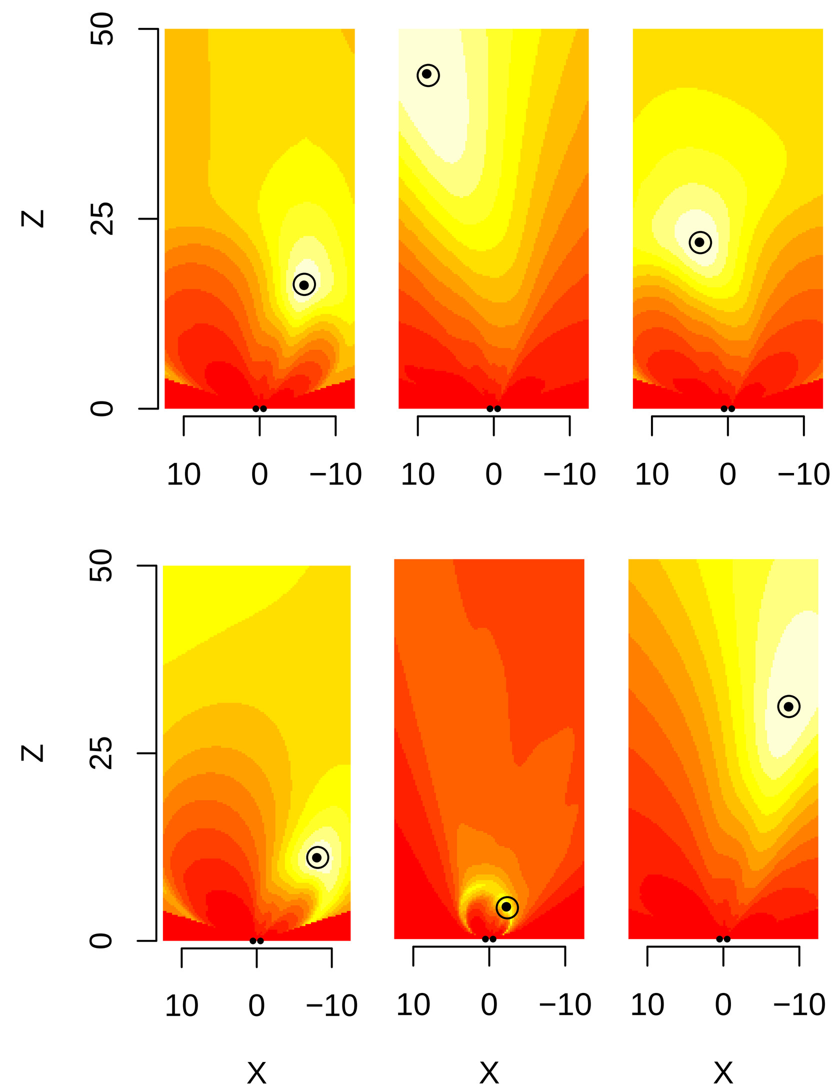


Figure 7: *Epipolar compatibility for six different fixations. The likelihood $\exp(-E(x,z))$ is plotted over the horizontal plane; the lighter-coloured regions are more probable. The eyes are shown at the bottom of each rectangle. Black dots mark the true fixations; unfilled circles are centred on the maximum value of the array.*

- Our results suggest that the error function E is smooth.
- The global minimum, over the present parameter range, identifies the true fixation-point.

9. Gaze Uncertainty Analysis

- We estimate the Hessian by fitting a quadratic basin to the minimum of the cost function $E(x,z)$. The trace of the inverse-Hessian was used as a measure of uncertainty, V.
- This was repeated for each of 50 fixations.
- Uncertainty increases rapidly with the range of the fixation-point, ρ.

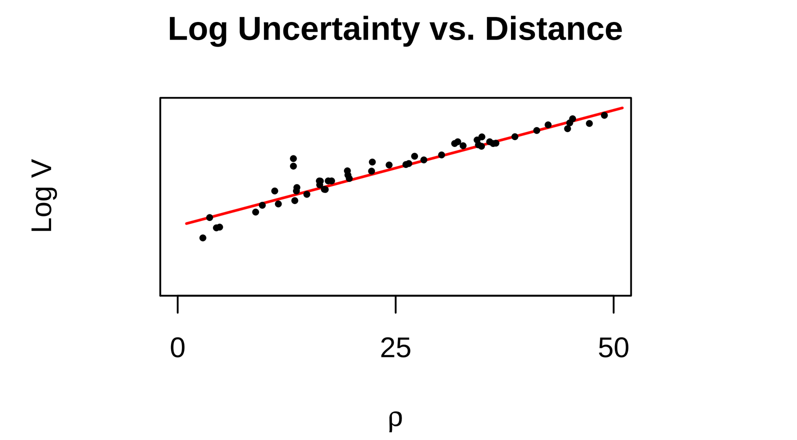


Figure 8: *Log-uncertainty of the gaze-estimate for 50 fixations. A linear fit is shown in red.*

- This means that it becomes unfeasible to estimate the range of distant fixation-points from geometric information.

10. Biological Relevance

Epipolar Transformations Each epipolar geometry defines a 'flow field' of disparity, cf. fig. 3. The analogous motion fields are known to be effective stimuli in area MST. We suggest that wide-field binocular extra-striate mechanisms may be tuned to epipolar transformations.

Gaze Representation The binocular gaze-point could be represented in a 2-D topographic map, cf. fig. 7. We have shown that the uncertainty of the gaze estimate increases as a function of fixation-distance. Hence we expect more neural resolution for nearby visual space.

11. Future Work

Theory We intend to analyze the structure of the cost function, and to investigate any geometric ambiguities.

Matching We will replace the matching procedure by a biologically plausible method, using phase-shifted binocular filters.

Control We plan to use a binocular robot-head to test a gaze-control strategy based on these ideas.

12. Acknowledgments

This work is part of the *Perception on Purpose* project, supported by EU grant 027268.