Biological
Cybernetics

Direct extraction of curvature-based metric shape from stereo by view-modulated receptive fields

A. J. Noest · R. van Ee · A. V. van den Berg

Abstract Any computation of metric surface structure from horizontal disparities depends on the viewing geometry, and analysing this dependence allows us to narrow down the choice of viable schemes. For example, all depth-based or slant-based schemes (i.e. nearly all existing models) are found to be unrealistically sensitive to natural errors in vergence. Curvature-based schemes avoid these problems and require only moderate, more robust view-dependent corrections to yield local object shape, without any depth coding. This fits the fact that humans are strikingly insensitive to global depth but accurate in discriminating surface curvature. The latter also excludes coding only affine structure. In view of new adaptation results, our goal becomes to *directly* extract retinotopic fields of metric surface curvatures (i.e. avoiding intermediate disparity curvature).

To find a robust neural realisation, we combine new exact analysis with basic neural and psychophysical constraints. Systematic, step-by-step 'design' leads to neural operators which employ a novel family of 'dynamic' receptive fields (RFs), tuned to specific (bi-)local disparity structure. The required RF family is dictated by the non-Euclidean geometry that we identify as inherent in cyclopean vision. The dynamic RF-subfield patterns are controlled via gain modulation by binocular vergence and version, and parameterised by a cell-specific tuning to slant. Our full characterisation of the neural operators invites a range of new neurophysiological tests. Regarding shape perception, the model inverts widely accepted interpretations: It *predicts* the various types of errors that have often been mistaken for evidence *against* metric shape extraction.

1 Introduction

Any extraction of metric object structure from (horizontal) binocular disparity must depend on the viewing geometry (e.g. von Helmholtz 1867; Mayhew and Longuet-Higgins 1982; Mayhew 1982), all the more so if one wants to extract metric *shape*, i.e. view-invariant surface curvature fields. Of course, no real system can be perfectly accurate, and, more importantly, the actual goal is to find a well-characterised, robust neural mechanism that offers a realistic operational redefinition of metric shape, i.e. one that not only adds some highly valued features (e.g. robust 'smoothing' and 'transparency') but also produces realistic 'errors'.[1] To present our approach and its neural implementation in the most accessible form, without losing the essential elements of the full metric surface-shape problem, we limit this paper

A. J. Noest (✉)
Functional Neurobiology Department, Utrecht University,
NEST, Limalaan 30, 3584-CL Utrecht, The Netherlands
e-mail: a.j.noest@bio.uu.nl

R. van Ee
Department Physics of Man, Utrecht University,
Princetonplein 5, 3584-CC Utrecht, The Netherlands

A. V. van den Berg
Functional Neurobiology Department, Utrecht University,
Padualaan 8, 3584-CH Utrecht, The Netherlands

---

[1] Note the essential step of inverting, right from the start, the usual interpretation of these errors as evidence against metric shape processing (e.g. Todd and Norman 2003). Our model actually *predicts* these errors.

to extracting only the sectional curvatures in epipolar planes, and we neglect eye torsion.[2]

The organisation of the paper reflects our gradual zooming in on a viable design. The rest of this introduction classifies existing approaches (Sect. 1.1), summarises hints from neurophysiology (Sect. 1.2.1) and psychophysics (Sect. 1.2.2), and reports new shape-adaptation experiments (Sect. 1.2.3) we performed to help us choose the appropriate neural shape code. In the body of the paper (Sect. 2) we first rule out existing approaches by robustness and geometric analysis (Sect. 2.1) and then start our stepwise design from a new exact expression relating object curvature to disparity structure and viewing parameters (Sect. 2.1.2). We derive a new family of receptive fields tuned to the relevant local disparity structure (Sect. 2.2) and use these to build the neural operators that directly extract object curvatures (Sect. 2.3). Many specific neural predictions ensue. In Sect. 3, we explain how our scheme inverts the usual interpretation of experiments that are often thought to argue *against* metric shape extraction.

1.1 A broad classification of existing approaches

Within the extensive literature (Howard and Rogers 2002, for review) one can distinguish a few general types of approach. Most of these have not yet been worked out to neurally explicit models, and for some this seems unfeasible. The same classification covers the sets of assumptions used (often implicitly) in most 'shape-from-stereo' experiments. One finds four main types of frameworks:

A. *'Classical', based on disparity/depth values* (von Helmholtz 1867; Mayhew and Longuet-Higgins 1982; Mayhew 1982; Gårding et al. 1995). Within this type of strategy, extracting explicit metric object shape from disparity would be a three-step transformation:

1. Correction of the horizontal disparity field for the viewing geometry, i.e. for the local slant and curvature ($1/R$) of the Vieth–Müller circle through the fixation point (Fig. 1). This stage produces the local depth (or its inverse, 'nearness') modulo an unknown linear transformation along the viewing direction ('affine' depth). This reduces families of shapes to 'reliefs'.
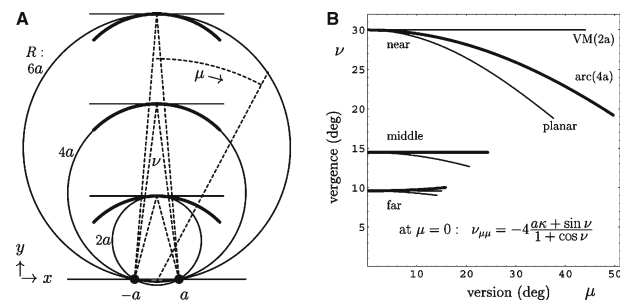
Fig. 1 A basic example of how changing the viewing distance affects the binocular disparity field evoked by a fixed-shape surface, here projected to a line segment (*thick*) on the plane of regard, with the eyes at $(x, y) = (\pm a, 0)$. **a** The same concave surface patch (here a circular arc of curvature $\kappa = 1/(4a)$), presented at three viewing distances, is more or equally or less curved than the three Vieth–Müller circles (zero disparity loci). Thus, its 'disparity curvature' $\nu_{\mu\mu}$ changes sign with distance. **b** The same stimuli (right half only) mapped into binocular coordinates (version μ, vergence ν). Note how the viewing distance affects the $\nu_{\mu\mu}$ of the stimulus image (*thick line segment*). The formula gives the exact result for $\mu = 0$; the general result, derived in Appendix A, will appear in Eq. (3). Both figures illustrate how the curvature $1/R$ of the Vieth–Müller circle leads to a deviation from the claimed (Rogers and Cagenello 1989) viewing-distance invariance of 'disparity curvature' $\nu_{\mu\mu}$: this approximation would only become perceptually acceptable (e.g. error below the known 5% Weber fraction) for patches with curvature radius less than 1/40th of the viewing distance, seen at slants less than 12° (Sect. 2.1.2 and Appendices A.3 and A.3.1)

2. Converting affine to metric depth (in a retinotopic frame) by a scaling that depends strongly on the fixation distance (measurable from the vergence ν and/or the structure of the vertical disparity field).

3. Computation of shape-in-depth from the field of metric depth values.

B. *'First-order', based on disparity gradient and/or surface slant* (Koenderink and van Doorn 1976; Lee 1999; Grossberg and Swaminathan 2004). Apart from using first-order (e.g. the full disparity-field deformation) rather than zero-order structure, this type of approach would still need three stages analogous to type A in order to arrive at metric shape. (So far, such models have extracted only affine shape, insofar as they computed any object, rather than disparity structure, at all).

C. *'Second-order', based on disparity or surface curvature* (Rogers and Cagenello 1989). As presented originally, this would be a one-step process: object shape (represented by a field of surface curvatures κ) would be extracted directly from the horizontal disparity field, simply by two-fold differentiation. Based on an approximate analysis, this 'disparity

---

[2] In Sect. 3, we sketch how the model can be completed. This requires no fundamentally new concepts or neural operators, but it would make all the formulae appear so much more complicated that it would obscure the really important notions, analyses and predictions that we present here.

curvature'[3] $v_{\mu\mu}$ (Fig. 1) was claimed to be invariant for fixation distance, at least at small versions and surface slants. Figure 1 illustrates that even this restricted invariance fails, unless the object radius of curvature ($1/\kappa$) is negligibly small compared to the viewing distance ($\approx 2R$), thus excluding many practical situations. However, we will show how the original mistake can be repaired, and the approach can be generalised to cover the full relevant range of viewing situations, while still allowing for a neurally plausible, one-step shape-extraction process.

D. *'Headcentric' disparity-/depth-based* (Erkelens and van Ee 1998). This proposal again implies a three-step transformation to compute shape-in-depth, but its operations and data representations differ fundamentally from those in a type A framework.

1. Coordinate transformation of each eye's image, separately, into a headcentric (HC) frame. Thus, the angular position of each eye is required at this stage.
2. Computing the (HC frame) 3D coordinates along surfaces from the pair of HC-frame images. This 'triangulation' requires knowing the interocular distance.
3. Computing shape-in-depth of the surface from its HC coordinates.

1.2 Experimental constraints and hints for model design

Some qualitative but crucial aspects of the model are suggested by existing or new experimental results. An important example is the choice of data representations. This entails two main questions:

(i) Which quantities encode perceived local shape-in-depth and its computational precursors? Note that much of the diversity of existing model frameworks (see above) boils down to their choice of extracting metric or affine depth, slant or curvature, from local disparity, or its first- or second-order structure.

(ii) Which coordinate frame carries the field of local codes? Whatever the local code at each stage, is it tied to a retinotopic, headcentric, or perhaps object-centred frame?

In collecting existing data relevant to the choice of representation, we also find some clues about the neural mechanisms which compute the codes, and we identify remaining ambiguities, which we then resolve by new experiments.

1.2.1 Neurophysiology: disparity tuning and gain modulation

Population coding of disparity distributions across space. It is well established (Ohzawa 1998; DeAngelis 2000) that disparity is extracted, from area V1 onwards, by neurons that are selective ('tuned') for specific pairs of locations on the retinae. The full selectivity of such a cell can be viewed as a single, 4D RF centred on a specific locus in the space spanned by the two dimensions of 'cyclopean retinal' location and two dimensions of disparity. In fact, most of these RFs lie in or near the 3D subspace of horizontal disparity tuning (Cumming 2002).

Gain modulation by vergence and version. The neural gain of many disparity-sensitive cells in V1 and V2 is known to be substantially modulated by vergence and/or gaze (Trotter et al. 1992, 1996; Trotter and Celebrini 1999), usually without significant shifting of the optimal disparity to which the cell is tuned. More precisely, the gain dependence on vergence v is generally monotonic, whereas its dependence on version μ is often 'one-humped'. Other groups found similar modulation effects driven by viewing geometry (Gonzalez and Perez 1998; Rosenbluth and Allman 2002). These results, especially the specific types of view dependence found by Trotter et al. (1992, 1996), provide useful hints and challenges for constructing a neurally and computationally viable model.

Coding of first-order disparity structure in areas MT and V4. Recent recordings in areas V4 (Hinkle and Connor 2002) and V5/MT (Nguyenkim and DeAngelis 2003) have found cells with selectivity for first-order disparity structure (gradients). For a substantial fraction of these cells, the selectivity appears to be actually based on the local surface slant (and tilt), which corresponds to disparity gradients that depend strongly on viewing geometry. Computing first-order structure is not required for extracting shape (second-order structure), but our computational design will identify a definite subsidiary role for slant (and, in 3D, tilt) selectivity.

Shape coding in area IT. Janssen et al. (1999, 2000, 2001) have found cells in parts of the IT cortex which respond specifically to curvature-in-depth of binocularly defined surfaces, with considerable position invariance laterally as well as in depth. The results leave open the question whether these cells are coding *object* or *disparity* curvatures, i.e. whether their output is properly corrected for viewing distance and slant. Nevertheless, comparing these cells to the gain-modulated disparity cells found in area V2 (Trotter et al. 1992, 1996; Trotter and Celebrini 1999) suggests that the V2 cells could feed into, or even be part of, a stage that we seek to model

---

[3] Throughout, we follow standard notation for partial derivatives, using letter subscripts. Table 1, at the head of Sect. 2, lists where all globally used symbols are introduced, developed, or illustrated.

here, i.e. a stage that extracts shape proper, but not yet with the position invariance (object-frame coding?) of the curvature-driven IT cells.

1.2.2 Existing psychophysics: direct extraction of object shape, based on retinotopic second-order disparity structure?

Realistic handling of most stimuli requires disparity-tuned coding. Some fundamental constraints and hints for modelling follow from qualitative, everyday visual abilities that are often taken for granted: any reasonably realistic handling of situations where visual data are partially missing, spatially interleaved, 'rough' or discontinuous, spiked with outliers, or (especially) transparent, requires replacing the usual model representations that limit disparity, depth, slant or shape to be *functions* of position by their mathematically and neurally natural generalisation: *blurred relations* (Noest 1994). Precisely this is what a disparity-tuned population code does. Even the relatively 'mild' visual input complexity in most psychophysical stimuli (also in our experiment, Sect. 1.2.3 and Appendix B) requires such a representation. At higher levels, joint tuning for disparity and its gradient appears (Lee 1999), as expected.

Blindness to depth per se, but precise perception of curvature. Classic, robust results strongly contradict depth-based coding and favour coding based primarily on curvature. For example, humans are simply blind to distance and disparity when no visual 'reference' objects are available (Erkelens and Collewijn 1985), and they suffer from Cornsweet-type depth illusions (Anstis et al. 1978). Even slant perception without reference is weak and error prone (as reviewed in Howard and Rogers 2002), although it stabilises with training (van Ee 2001). Large errors also occur in many (but not all) depth tasks using curved surfaces [Todd and Norman (2003) give a good overview, but our model will invert their conclusion]. In striking contrast with depth or slant, the curvature-in-depth of surfaces is usually perceived with good precision – down to 5% Weber fraction (Rogers and Cagenello 1989).

These contrasting results plead strongly against schemes in which a depth-/slant-based code is computed before shape is extracted from it (types A, B or D above); instead, they suggest direct extraction of curvatures (as in type C). This also fits the improved detectability of small local depth perturbations when referenced to a smooth surface (Glennerster et al. 2002).

Adaptation hints at retinotopy and shows direct object-shape extraction. Adaptation experiments can offer a relatively uncluttered view of how shape-in-depth is represented neurally,[4] but one must allow for superimposed adaptation contributions from all processing stages that precede or influence the object shape representation we seek. Indeed, we already know from neurophysiology that at least the *input* to the shape-extraction process is represented by a retinotopic disparity-tuned code.

The first study of 3D after-effects (Köhler and Emery 1947) already concluded that binocularly defined shape is also coded retinotopically, but this conclusion would be unsafe if the overall after-effect was dominated by contributions from the binocular input stage. Over the next four decades, a series of adaptation studies (reviewed in Howard and Rogers 2002, and specifically Ryan and Gillam 1993 and references therein) gradually managed to separate out the input-stage contribution (which has very precise retinotopy), leaving other contributions that proved to be also retinotopic but less precisely so (e.g. 'phase independent'). However, these experiments focused not on *shape* but on depth or slant-/tilt-related disparity gradients (Ryan and Gillam 1993 and references therein).

Recent experiments (Berends et al. 2005) have revealed that the representation of disparity-defined slant is corrected for viewing geometry. However, independently of whether this encodes the perceived slant/tilt (Knapen and van Ee 2006), none of these results answer our question, since 'shape' is by definition a slant-invariant quantity – it is natural (and more robust, as we show later) to compute it directly, instead of from slant/tilt fields,[5] and the higher accuracy of curvature perception mentioned above strongly hints at such a scheme. In all, existing experiments suggest, but do not determine, that coding of *shape* is also retinotopic. Resolving this uncertainty is one of the motivations behind our adaptation experiments (Sect. 1.2.3 and Appendix B).

A highly relevant set of recent shape-adaptation experiments (Domini et al. 2001; Duke and Wilcox 2003) indicates that shape extraction from disparity immediately yields *object* shape. For example, Domini et al.

---

[4] Adaptation can avoid many of the complications that arise in interpreting measurements of JNDs or systematic errors (bias) in perceived depth, slant or 'shape': JNDs depend strongly on all the unknown neural noise levels, and biases reflect the unknown errors in viewing geometry estimates that *any* model which aims to extract object shape must use. Indeed, our model ultimately provides a new explanation for many known perceptual errors, in a way that actually inverts existing interpretations (Sect. 3).

[5] Note that this principle does not conflict with the slant tuning which will emerge (e.g. Sect. 2.2.3) as an indispensable feature of the neurons which directly extract object curvature. Their *implicit* slant encoding controls shape corrections and weakly constrains the perceptual slant ambiguity inherent in any true shape encoding.

(2001) had subjects adapt to surfaces of various curvatures κ at various viewing distances (and thus vergences ν) so as to dissociate the role of 'disparity curvature' $\nu_{\mu\mu}$ from that of object curvature κ. A simple illustration of the distinction is shown in our Fig. 1, and it is quantified by the relation $\nu_{\mu\mu} = -4\frac{a\kappa + \sin\nu}{1 + \cos\nu}$, which is exact for the simple 'straight-ahead, frontal' viewing geometry used in these experiments. Note that the ν-range used was indeed sufficient to invert the sign of $\nu_{\mu\mu}$ (as in our Fig. 1) in some of the conditions used. The conclusion that *object* instead of disparity curvature is the basis for the shape code is a vital constraint on any neurally explicit model: the machinery that extracts the shape code from a disparity structure must correct it *in the same step* for the effect of viewing distance. Their data also reinforce previous hints (see above) that the shape code is curvature based, but it does not yet exclude the possibility that e.g. a slant-based code might serve as implicit shape code (or as computational precursor of an explicit code), as long as such a slant field would also be properly corrected for the effect of viewing distance (again referring to Fig. 1, one may imagine the required rotation of local tangents to the (μ, ν)-images of the stimuli).

Very similar conclusions can be drawn from data obtained by Duke and Wilcox (2003), who used a similar experiment but used shapes with *illusory* curvatures induced by manipulating the vertical disparity field, so as to correspond to a different viewing distance. Besides confirming the previous conclusions on correction of shape for viewing distance, this experiment shows functional equivalence between various signals (vergence and vertical disparity structure) which normally contain viewing distance information.

The remaining uncertainty about whether the relevant shape code is dominantly *curvature* based prompted us to include this question in our adaptation experiments.

1.2.3 New adaptation experiment: retinotopic curvature-based shape

The full description of the design, methods, results and interpretation of our experiment is given in Appendix B, but the main points are as follows.

The experiment aimed to test whether perceived shape is coded by curvature tied to a retinotopic frame. As possible alternatives, we considered shape representations based on local disparity or its gradient, coded in either a retinotopic or a HC frame. Such schemes could coexist, giving rise to multiple contributions to the overall strength of adaptation K, which we quantify

as the curvature of a test stimulus adjusted to appear flat after adaptation to a fixed stimulus (Appendix B.1). What matters to our model is the difference ΔK between adaptation strengths K_+, K_- measured after adapting to concave, respectively convex, versions of otherwise identical adaptation stimuli. Indeed, ΔK separates mechanistically interesting adaptation-dependent K-contributions from the well-known bias in perceived curvature which we attribute to errors external to our model.[6]

To allow separating the retinotopic and HC contributions, the adaptation and test stimuli are viewed either in the same or opposite HC hemispheres. To separate curvature coding from lower-order coding of shape, we use either 'static' central fixation or 'moving' fixation, where one tracks a marker that traverses the width and depth of the adaptation stimulus (Appendix B.1).

In summary, we use $2 \times 2 \times 2 \times 2$ conditions:

– Two adaptation shapes: convex/concave, with fixed curvatures $\kappa = \pm\kappa_0 = \pm\frac{4}{3}\,\mathrm{m}^{-1}$
– Two types of fixation during adaptation: 'static' or 'moving' laterally and in depth
– 2×2 combinations of Right/Left adaptation and testing (at $\mu = \pm 15°$)

HC shape coding predicts that adaptation does not transfer between opposite hemispheres, whereas retinotopic shape coding should transfer perfectly. Thus, we collect into one ('transfer') group all ΔK data from conditions with a (adapt→test) sequence of the type (L→R) or (R→L). Likewise, all data from (L→L) and (R→R) conditions go into a 'no-transfer' group. Combined with the 'static fixation' versus 'moving fixation' dimension, this leaves four groups of ΔK data (Appendix B.2).

All the conclusions of interest to our model construction can now be read simply from a 2D scatterplot (Fig. 2) of the means $\langle\Delta K\rangle$ within each of the four groups: separately for the 'static fixation' and the 'moving fixation' data, each pair of $\langle\Delta K\rangle$ for the 'transfer' and 'no-transfer' groups is used as the x, y-coordinates of a symbol in the scatterplot, using ellipses for 'moving fixation' and crosses for 'static fixation' data, with their widths along each axis representing the SEM values.

The patterns in this data summary (Appendices B.2 and B.3 for detailed data and interpretations) fully

---

[6] This bias has contributed to many experimental results, going back at least to the classic 'horopter' problem (von Helmholtz 1867). As explained in Sect. B.3, it is the expected effect of errors in an external control signal (related to fixation distance) required by any model. Our model minimises the impact of such errors relative to existing designs. The unavoidable remaining bias does not convey information about the processing stages we model in this paper.

Fig. 2 Adaptation strength measures ($\langle \Delta K \rangle / (2\kappa_0)$, see text for details) for interhemispheric 'transfer' conditions '(L→R) & (R→L)' versus those for 'no-transfer' conditions '(L→L) & (R→R)'. *Left* using cylindrical patch stimuli. *Right* using spherical stimuli. Per panel, the symbol *colours* distinguish data from different subjects
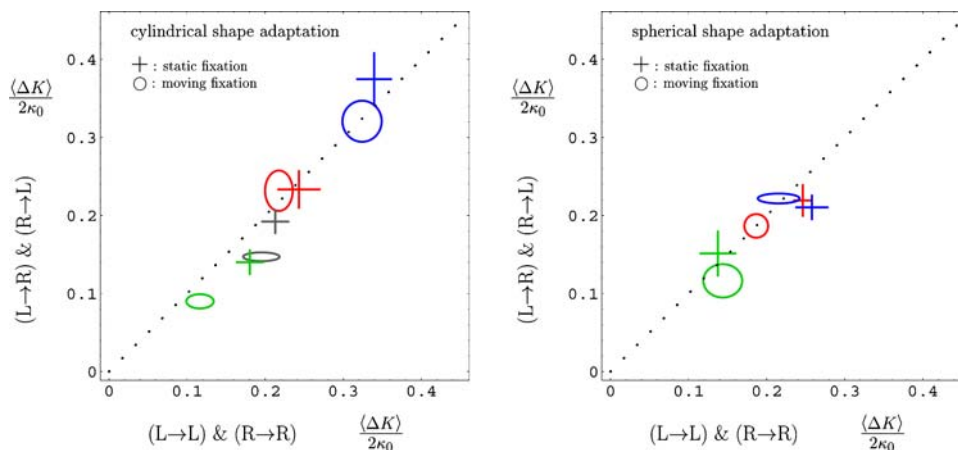
confirm the retinotopic, curvature-based nature of shape coding:

1. All symbols lie close to the diagonal, as predicted by any retinotopic code (whatever its local type). They lie many SEM units from their projections onto the horizontal axis, the locus predicted by HC coding.
2. Per subject, the ellipse (moving fixation data) is generally much nearer to the cross (static fixation data) than to the origin. Idealised curvature-based shape coding predicts these pairs of symbols to coincide, whereas the circles only should fall on the origin for a shape code using disparity *values* or their gradient. Thus, our data favour at least a dominant contribution from curvature-based shape coding, consistent with other results (Domini et al. 2001; Duke and Wilcox 2003). Moreover, realistic neural models of curvature-based shape are expected to produce side effects of the same type as the minor deviations from the ideal we observed (Appendix B.3).
3. A subject-dependent scale factor represents most of the variance in the adaptation strengths. This is irrelevant for the model-design decisions we must make.

2 Neural model: extracting local object curvature

Our first step must be an analysis of the geometric structure and computational demands of the task. The construction of neural model(s) then unites the theoretical results with the neurophysiological and psychophysical hints and constraints. In fact, our neural model does not merely implement the results of our initial formal analysis, but actually generalises them – we start from a widely familiar setting and gradually extend it, indicating where, how and why several less familiar features arise in our model(s). A few advanced notions are eventually reached, but the neural stages that implement

Table 1 References to sections, figures or equations where symbols (except those used only locally) are introduced, significantly developed, or illustrated

Symbols	Sections, figures, equations
$A, A(\mu, \nu)$	(2.2.1)
$A_{\nu\nu}$	(2.2.3)
a	Fig. 1, (2.1), (A.1)
$B(u, v, \theta)$	(2.3.2), Fig. 9
$C_{\pm}$	Eq. (13)
D	(2.1.1), (A.1)
$G, G_{uuv}, G_{\nu\nu}$	(2.2.2), (2.2.3), Fig. 3, (2.3.1)
$K, \Delta K$	(1.2.3), (B.1), (B.2)
$N_{\pm}$	Eq. (11)
R	Fig. 1, (A.1)
$S(\mu, v, \tau)$	(2.2.3), Fig. 4, (2.3.2)
u, v	(2.2.3), Fig. 3, (2.3.1)
W_0, W_1	(2.1.2), (2.3.1), Fig. 7, Fig. 8, (2.3.2), (A.3.2)
x, y	Fig. 1, (2.1), (A.1)
α	(2.1.1), Eq. (18)
β	(2.1.1), Eq. (2), (2.1.2), Eq. (18)
γ	(2.2.2), Fig. 5
δ, ϵ	(2.1)
θ	(2.3.2), Fig. 9
κ	Fig. 1, (2.1.2), Eq. (3), (2.3.1), (2.3.2), (A.3)
μ, ν	Fig. 1, (2.1), (A.1)
μ_0, ν_0	(2.1), (2.1.1)
$\nu_\mu, \nu_{\mu\mu}$	(2.1.2), (2.2.2)
ρ	(2.3.1), Fig. 6, (2.3.2), Fig. 9
$\sigma_\delta, \sigma_\epsilon, \sigma_\tau$	(2.2.1), (2.2.3)
$\tau, \hat{\tau}$	(2.2.2), (2.2.3), (2.3.1), Fig. 8, (2.3.2)

them may also serve as a simple, concrete illustration of the underlying mathematics.

2.1 Mathematical basis: geometric structure and robustness

To compute the surface curvature that defines local metric object shape it is neither necessary nor sufficient

to compute depth and/or slant *per se*. To find what is required, we first explore the major consequences of choosing the appropriate geometric setting and demanding computational robustness. New analytical results (Appendix A) then guide us to the appropriate neural operators.

We start by considering planar curves, which are the intersections of object surfaces with an epipolar (constant elevation) half-plane, equipped with Cartesian coordinates (x, y), as well as bicentric coordinates (μ, ν), version and (full) vergence (Fig. 1), bounded by $2|\mu| + \nu < \pi$ and $\nu \geq 0$. We do not treat the full 3D problem in this paper, but note that the analysis is identical in any epipolar plane.

It is usual to identify this $(x, y > 0)$-plane with a 'plane of regard' through eye centres (at $x = \pm a, y = 0$) and a 'fixation point' at version and vergence angles μ_0, ν_0. The (azimuthal) 'cyclopean retina' coordinates are then the disparity $\delta = \nu - \nu_0$ and eccentricity $\epsilon = \mu - \mu_0$, both of which are accurately coded in the visual system,[7]

By contrast, the usual 'fixation vergence' ν_0 cannot be a well-defined quantity. The problem is not just that the formal 'fixation point' is generically *non-existent*, but that *large relative errors* must occur on the small angular difference between the (3D) gaze directions of the left and right eye: at practical viewing distances, the gaze difference is one to two orders of magnitude smaller than the range available to each gaze, making it very sensitive to left-/right-uncorrelated gaze errors (Steinman et al. 1982).

The error component relevant to our analysis can be captured within a 2D setting: having chosen to process elevation disparity in a separate pathway (irrespective of its results feeding our model with viewing geometry data), we may redefine a (generic) 'plane of regard' with its elevation midway between that of the left- and right-eye gaze, project each gaze on it, and redefine μ_0 and ν_0 as the mean and difference angles of the projected left- and right-gaze directions (Helmholtz azimuths). Note: ν_0 can now be negative.

More generally, all disparity-processing schemes must face the fact that 'viewing distance' is unavoidably noisy, and may even be undefined. In analysing the computational robustness to such uncertainty, we use the redefined ν_0 to represent the *combined* viewing distance information (possibly regionally different) used by disparity processing.

---

[7] Without eye torsion, they would be limited only by the resolution of pairs of ('hardwired') retinal-frame local signs. Correcting for realistic eye torsions, even when these are not very accurately available, is possible without spoiling the δ, ϵ accuracy we need. In this paper, we simply assume this done.

2.1.1 Viable computations must be robust to fixation vergence error

Existing approaches differ widely in how they deal with the ν_0-uncertainty (and possible zero-crossing). One extreme is to simply calculate depths using whatever ν_0 estimate is available. This occurs in e.g. 'disparity scaling or normalisation', which is often implicitly assumed in interpreting psychophysical data; see below for the implications. The opposite extreme is to retreat from metric to affine shape ('relief'), thus leaving unknown a scaling along the ν (or ν_0)-direction. This occurs in many different schemes, e.g. in 'disparity correction' and in deformation- or angular-disparity-based methods. There are few proposals (e.g. 'disparity curvature') for ν_0-robust extraction of metric object geometry, and so far these are valid only under severe restrictions (Appendix A.3).

We rank and select metric-shape computations by their degree of sensitivity to ν_0 (and less critically to μ_0 and slant). This criterion makes sense since disparity-driven shape percepts do not generally show anything like the extreme shape distortions predicted by most existing models (see below). Our robustness ranking may be viewed as a graded version of seeking full 'iseikonic' (μ_0, ν_0)-invariance (Koenderink 1992).

Zeroth-order structure: depth (or depth differences) from disparity? Metric type A models essentially try to recover cyclopean distance $D = 2a \cos \mu / \sin \nu$ from disparity δ. Besides being at best only a first step towards computing metric shape, this is not a viable method because division by the small term $\sin \nu \approx \nu_0 + \delta$ amplifies ν_0-errors into large D-errors. Moreover, it predicts that fusing a pair of slowly outward-moving stereo images should drastically distort the distant parts of a perceived surface towards infinity (as $\nu \downarrow 0$), followed by an absurdity: $D < 0$ when $\nu < 0$. Not even the lesser of these distortions is seen (thus also excluding that the system clips ν_0 or ν).

Similar breakdowns preclude the use of disparity differences $\Delta \delta$ (often called 'relative' disparity) to recover depth differences ΔD: The $\Delta \delta$ are free of the ν_0 uncertainty, but the problem resurfaces in computing the depth differences

$$\Delta D \approx -2a \left(\Delta \delta \cos \mu \frac{\cos \nu}{\sin^2 \nu} + \Delta \varepsilon \frac{\sin \mu}{\sin \nu} \right). \quad (1)$$

The division by the very small term $\sin^2 \nu \approx (\nu_0 + \delta)^2$ actually makes this step ('disparity scaling or normalisation') an order of magnitude more sensitive to ν_0-errors than direct D recovery, for $\nu > 0$. For parallel viewing ($\nu_0 = 0$) of stereo images, disparity scaling predicts wildly blown-up and 'folded' depth structure, contrary

to observation. For $\nu > 0$, the two schemes are equivalent when judged by the same measure, e.g. ΔD.

First-order structure: surface slants/slopes from disparity gradients? The gradient ν_μ, accurately measurable as the disparity gradient δ_ε, is obviously (ν_0, μ_0)-invariant and thus blind to depth per se, but it carries some slant information. Computing e.g. the usual HC slant $\alpha \equiv -\arctan y_x$, or 'slope' $-y_x$, is marred by an unbounded μ_0-sensitivity (Appendix A.2). This is avoided only when computing the 'cyclopean direction-centric' slant $\beta = \alpha - \mu$, or its slope

$$\tan \beta = \tan \mu + \nu_\mu / \tan \nu. \qquad (2)$$

The division by $\tan \nu \approx \nu_0 + \delta$ means that computing $\tan \beta$ is only degree-1 ν_0-sensitive,[8] but even knowing the true $\tan \beta$ as a function of μ determines only a 'relief'. To recover metric depth differences would at least require scaling by viewing distance, equivalent to an extra division by ν_0, which gives this (and any other first-order) scheme the same degree-2 error sensitivity as all zeroth-order methods.

Nevertheless, the $\tan \beta$ field itself will play an important but subsidiary role in our neural schemes for curvature extraction. For this it is enough that the $\tan \beta$ errors occur mainly where $|\beta| \approx \pi/2$, i.e. usually only in a visually narrow strip near contours (where monocular cues are more informative). Robustness in terms of the slant angle β is formally better there, but this is not the visually relevant measure when $|\beta| \approx \pi/2$.

2.1.2 Object curvature from disparity differentials

Given our goal of extracting local object curvature κ, it is natural to try to compute it from the second differential $\nu_{\mu\mu}$, which is obviously free of (ν_0, μ_0) errors, and is expected to be accurately measurable as $\delta_{\epsilon\epsilon}$. Before any metric-level analysis, we must address a more fundamental issue:

Visual geometry differs structurally (i.e. not just by choice of coordinates) from Euclidean geometry, essentially in that it gives one of its dimensions a fixed, special direction and orientation:[9] In our case, a fixed, signed ν-axis is imposed by the special role of cyclopean visual rays. This seemingly trivial fact has many profound consequences throughout our computational design, its

neural implementation and its predictions. The most basic consequence is that each local curvature κ carries an *extrinsic* sign ('convex/concave'), defined only for $|\beta| \neq \pi/2$, whereas in Euclidean geometry the κ-sign (once fixed) is invariant under object rotations. Note that Euclidean rotations eventually violate visual curvature continuity (even for 'transparent' objects) by causing a 'sign flip' in κ. To be perceptually relevant, our model should extract visual-sign curvatures κ, with $|\kappa|$ matching the magnitude of the Euclidean curvature as far as possible (cf. Appendix A). Clearly, this must fail in some way near the sign-flip discontinuities. Thus, we have identified a second robustness aspect: avoiding perceptually serious errors where $|\beta| \approx \pi/2$, i.e. where the local slope $|\tan \beta|$ diverges and the visual curvature κ changes sign discontinuously.

Visual object curvature κ in terms of $\nu_{\mu\mu}$, ν_μ and viewing geometry parameters. The original estimate (Rogers and Cagenello 1989) was that κ and $\nu_{\mu\mu}$ would be simply proportional, independently of viewing distance (and perhaps, of version and slant). If true, computing object curvatures would be fully robust to viewing geometry and accuracy-limited only by extracting $\nu_{\mu\mu}$ from the neural disparity encoding. Later analysis (Koenderink 1992) identified that κ actually depends also on ν_μ and all viewing geometry parameters. Here, we give the exact result in closed form and derive those 'corrections' to $\nu_{\mu\mu}$ which are required to satisfy the known perceptual accuracy and neural viability constraints.

In our setting, and choosing κ to be positive for visually 'convex' shapes, exact expressions for κ can be found by applying the classical formalism for parametric planar curves [Appendix A, leading up to Eq. (24)]. The full result can be reduced to the form

$$
\begin{aligned}
-a\kappa = & \left\{ \frac{1}{4} \nu_{\mu\mu} (\cos 2\mu + \cos \nu) \right. \\
& \left. + \left(\sin \nu + \frac{1}{2} \nu_\mu \sin 2\mu \right) \left(1 - \frac{1}{4} \nu_\mu^2 \right) \right\} \\
& \times \left\{ 1 + \frac{\nu_\mu \sin 2\mu}{\sin \nu} \right. \\
& \left. + \left(\frac{\nu_\mu}{2 \sin \nu} \right)^2 \left[(\cos 2\mu + \cos \nu)^2 + \sin^2 2\mu \right] \right\}^{-\frac{3}{2}}.
\end{aligned}
$$
(3)

At first glance, Eq. (3) might appear unsuitable for neural implementation, but a bit more scrutiny of its computational structure and robustness opens up the route to realistic neural models which even generalise the computation well beyond the representational limitations of Eq. (3).

First, note that Eq. (3) is just a *linear* expression in $\nu_{\mu\mu}$, but with coefficients depending on all other

---

[8] The formal μ_0-error sensitivity due to the term $\tan \mu = \tan(\mu_0 + \varepsilon)$ is negligible compared to the ν_0-sensitivity: eye mechanics and optics constrain the binocular visual field so that μ cannot reach the small neighbourhood of $\pm\pi/2$ where μ_0-sensitivity would arise.

[9] The mathematical meanings of these terms are interchanged in neurobiology, and sometimes in psychophysics. This does not matter here. Later on, the local context clarifies the relevant meaning.

parameters. It will prove useful to group these in either of two ways: writing Eq. (3) as

$$- a\kappa = (b\nu_{\mu\mu} + c)/d = W_1 \nu_{\mu\mu} + W_0 \qquad (4)$$

defines the two coefficient sets as (b, c, d) and (W_0, W_1). The original proposal (Rogers and Cagenello 1989) corresponds to neglecting $W_0 = c/d$ and replacing $W_1 = b/d$ by a constant. This simple approximation is insufficient because its errors exceed the 5% Weber fractions they measured, except in an unrealistically restricted regime where the curvature radius is less than 1/40th of the viewing distance while the slant is less is than 12° (Appendices A.3 and A.3.1); moreover, it is refuted psychophysically by the Domini et al. (Domini) experiment.

The factor $b = (\cos 2\mu + \cos \nu)/4$ is innocuous since it is smooth and positive for any practical viewing geometry, including $\nu < 0$. The additive c accounts for e.g. the distance-dependent curvature-adaptation data (Domini et al. 2001): its term $\sin \nu$ reflects the required correction for the curvature of the Vieth–Müller circle (radius $R = a/\sin \nu$). Both other terms in c display the special role of $|\nu_\mu| = 2$. Note that this special value arises here independently of object opacity (as long as a surface patch is visible at all). Indeed, it pertains as well to 'transparent' or 'sampled' surfaces, when a local surface tangent becomes opposite-sense mapped on the two retinae. In any case, such large $|\nu_\mu|$ occur only close to a sign-flip discontinuity in κ (and for non-matched binocular 'noise'), so little is lost and robustness is gained if we design our system to avoid processing any data in this regime (see below).

The denominator d requires more scrutiny. Indeed, its terms $\nu_\mu/\sin \nu$ seem to destroy robustness, giving d an even worse (asymptotically degree-3) sensitivity to ν_0 errors. However, as in the previous cases, what matters is whether the blow-up of ν_0-errors (or any means of 'clipping' them, or locally avoiding computation of curvature at all) causes perceptually unrealistic distortion, instability or undefined status of object shapes. This does not occur here, as can be seen as follows.

For simplicity, consider the 'far-field' (small ν) regime.[10] This yields (Sect. A.3.1)

$$d = \cos^3 \mu \left\{ 1 + \left(\frac{\nu_\mu}{\nu_0 + \delta} + \tan \mu \right)^2 \right\}^{3/2} + \mathcal{O}(\nu^2)$$

$$= \left(\frac{\cos \mu}{\cos \beta} \right)^3 + \mathcal{O}(\nu^2). \qquad (5)$$

---

[10] This first-order approximation in ν is accurate under natural viewing conditions, and no drastic new effects occur until one reaches the unrealistically close-up $R \approx a$ regime where $\nu \approx \pi/2$.

These two forms are mutually related via the slope Eq. (2), up to $\mathcal{O}(\nu^2)$. Note that the first form, which follows directly from Eq. (3), shows explicitly the dependence on the neurally measurable parameters. Both forms imply that none of the mentioned catastrophic shape distortions will occur, even without any measures to avoid local blow-up of d at zero-crossings of $\nu_0 + \delta$. Firstly, the local $\kappa \propto 1/d$ will go to zero, i.e. the distortion consists only of the shape being approximated locally by its tangent plane. Secondly, it is clear especially from the second form that the ν_0-sensitivity is concentrated where the true slant $|\beta| \approx \pi/2$, i.e. near the extrinsic sign flips of κ, where *any* method must fail. Moreover, one can now reduce the already moderate perceptual errors further by measures that had been hitherto excluded because they would just replace one drastic distortion by another. For example, one can now safely clip ν at a positive value corresponding to the distance at which the finite neural disparity resolution already wipes out any depth structure; and one can simply *avoid* computing a local κ when $|\beta| \approx \pi/2$, i.e. where any result is known a priori to be unreliable. In that case, the shape of smooth surfaces remains constrained by the κ-field outside these loci.

2.2 Neural basis: RFs for extracting local disparity structure

2.2.1 Data representation: disparity-tuned cell activity

The disparity input to our neural computation of metric curvature consists of a field of activities $A(\epsilon, \delta) \geq 0$ of neurons which are tuned jointly for fixed disparity $\delta = \nu - \nu_0$ and cyclopean retinal location $\epsilon = \mu - \mu_0$ in the (extended) plane of regard (Sect. 2.1). The (say, Gaussian) RFs that define $A(\epsilon, \delta)$ can have a variety of widths $\sigma_\epsilon, \sigma_\delta$, but since we hardly use the across-scale structure which this provides, we drop the σ indices until needed. Similarly, we usually abbreviate $A(\epsilon, \delta)$ as A, or even write it as $A(\mu, \nu)$, a helpful abuse of notation motivated by the (μ, ν) dependence of the κ-computation [Eq. (3)]; no actual transformation into the HC frame (μ, ν) is implied or required.

Using the A-field replaces the conventional but highly restrictive use of disparity *functions* $\nu(\mu)$: fields $A(\mu, \nu)$ represent blurred *relations* (Noest 1994) between μ and ν. Our analytical results [Eq. (3), etc.] were derived in the $\nu(\mu)$-setting, so we should generalise all κ-computations to the A-setting, which is not just mathematically natural but also neurally and perceptually much more suitable and robust (see Sect. 1.2.2 and below). Our construction of the neural model essentially *is* that generalisation of Eq. (3). The original expressions serve as a
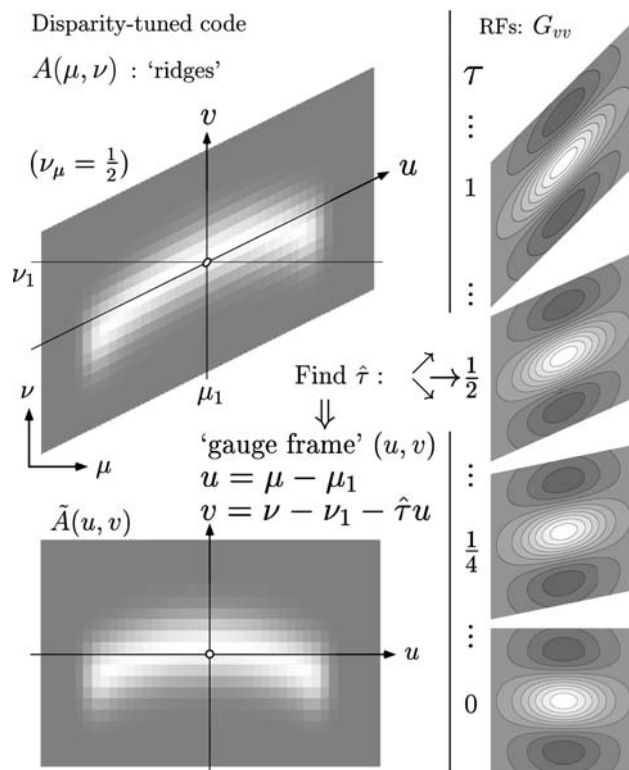
Fig. 3 Determination of local first-order contact τ of A-'ridges' (*top left*), the neural generalisation of formal gradient ν_μ of a disparity function $\nu(\mu)$. The local RF-family $G_{\nu\nu}$ (*right*) forms a τ-tuned code which can be 'read out' by e.g. 'winner take all'. Each τ defines a local "gauge frame" (u, v). *Bottom left* A-field is remapped into the (central) gauge frame, illustrating that such frames isolate the second-order contact γ (no actual A-remapping occurs)

simple but useful guide through the construction. In this paper, we use only part of the representational richness offered by the A-field coding, but what we use is crucial to most of what the model explains or predicts.

'Ridges' in A encode (and generalise) the formal 'curves'. The A-field automatically generalises the notion of 'surface' (here, 'curve') to a large variety of stimuli that generate activity 'ridges':[11] For curves in the old sense, the ridge essentially follows the graph of their disparity function $\nu(\mu)$. A finite-length example of such a ridge, with a blur scale set by the RF widths, is shown on the top left-hand side of Fig. 3. Note that very similar ridges will be created by curves with small-scale noise, and by sets of dot stimuli that sample an implicit curve – the blurring effectively 'smoothes' noise and 'connects the dots', up to its width scales $\sigma_\varepsilon, \sigma_\delta$. This regularised

---

[11] 'Fluffy' objects (e.g. a bush, furry animal, tree crown, etc.) generate ν-asymmetric 'scarps' rather than 'ridges' at their 'surface'. This requires extending all our local operators below to combinations of RFs of odd/even ν-symmetry. For the sake of readability, we do not write this out throughout the design.

encoding of object surfaces (Noest 1994) changes the data format used in computing κ: the role of the formal derivatives $(\nu_\mu, \nu_{\mu\mu})$ is taken over by the local geometry of A-ridges.

2.2.2 The Galilean (μ, ν)-geometry dictates novel RF-families

At first glance, it may seem that extracting the required A-ridge features is identical to the well-known neural extraction of line-like features in monocular vision by families of rotated Gaussian-derivative RFs (e.g. Koenderink and van Doorn 1987). But note: the special status of the ν-direction warns us that the very notion of 'geometric structure' in our (μ, ν)-plane must differ fundamentally from the familiar Euclidean notion. Indeed, the (μ, ν)-geometry is Galilean, not Euclidean (Appendix A.2. Thus, we must invent new RF families in (μ, ν)-space that respect the Galilean structure of its geometry.

One relevant consequence of the anisotropy of the (μ, ν)-geometry is that the Euclidean notion of rotation (incrementing angles) is replaced by the notion of incrementing slopes: the coordinate transformation for 'rotation' by τ [say, around $(0, 0)$] becomes $(\mu', \nu') = (\mu, \nu + \tau\mu)$. Clearly, this satisfies the basic demand of leaving any line parallel to the ν-axis invariant. Rigidly 'rotating' the graph of a function $\nu(\mu)$ thus simply adds τ to all its gradients ν_μ, and extending this to non-rigid local 'rotations' leads to the fact that the (μ, ν)-geometry analogue of 'curvature' (γ, say) simply yields $\nu_{\mu\mu}$ when evaluated on $\nu(\mu)$. Note the simplification compared to the familiar result $\nu_{\mu\mu}/(1+\nu_\mu^2)^{3/2}$ in Euclidean geometry. As a consequence, the first stage of generalising Eq. (3) is strikingly simple: we can just replace ν_μ and $\nu_{\mu\mu}$ by τ and γ, the local 'first- and second-order contact' parameters.

Now that we know the appropriate notion of 'rotation', generating the RFs for extracting the local geometry parameters of any order is straightforward. For $\tau = 0$, taking (μ, ν)-derivatives has the same meaning in both geometries, hence we can simply use the familiar Gaussian-derivative RFs of the Cartesian type: $\partial_\mu^n \partial_\nu^m G(\mu, \nu)$. The τ-parameterised RF-family belonging to any particular $\tau = 0$ exemplar then follows by 'rotating' its RF, which yields $\partial_\mu^n \partial_\nu^m G(\mu, \nu - \tau\mu)$.

2.2.3 Finding ridges and their first-order contact τ: the $A_{\nu\nu}$- and S-fields

The first type of operator we need is one that finds A-ridges and their local first-order contact parameter τ (Fig. 3).

At $\tau = 0$, the required RF is simply $G_{\nu\nu} \equiv \partial_\nu^2 G(\mu, \nu)$, well known in the Euclidean setting (Koenderink and van Doorn 1987). In our Galilean geometry, the τ-family of this exemplar, centred at a point (μ_1, ν_1), becomes

$$G_{\nu\nu} \equiv G_{\nu\nu}(u, \nu);$$
$$(u, \nu) = (\mu - \mu_1, \nu - \nu_1 - \tau(\mu - \mu_1)), \qquad (6)$$

where (u, ν) is the local 'gauge frame' (Koenderink and Richards 1988 and see below). In the shorthand $G_{\nu\nu}$, the ν implicitly contains the dependence on τ, which will prove to be a dynamic parameter in our model. Figure 3 shows some examples of $-G_{\nu\nu}$ RFs and of a gauge frame belonging to one of them. Ridge-finder neurons indeed need the sign-inverse RFs $(-G_{\nu\nu})$, since an object surface generates a 'ridge', not a 'valley'. We denote the response of such cells by

$$A_{\nu\nu} \equiv [-G_{\nu\nu} \star A(\mu, \nu)]^+;$$
$$[x > 0]^+ = x, [x \leq 0]^+ = 0, \qquad (7)$$

where $\star$ denotes convolution and $[x]^+$ the usual 'half-wave rectification'.

Note that the whole collection of $A_{\nu\nu}$ represents stimulus 'curves' as a neural activity field defined on a 3D (μ, ν, τ) parameter space: any particular $A_{\nu\nu}$-cell has a simple (one-hump) tuning profile centred on a fixed τ and cyclopean retinal locus (ϵ, δ). The τ-tuning curve (centred at τ_1, say) is $\{1 + [\sigma_\epsilon(\tau - \tau_1)/\sigma_\delta]^2\}^{-1}$. Apart from normalisation, this is a Cauchy distribution of width $\sigma_\tau = \sigma_\delta/\sigma_\epsilon$.

The fact that $A_{\nu\nu}$-responses are jointly tuned for zeroth- and first-order disparity structures [δ and $\tau \approx \nu_\mu$, both of which are parameters in the κ expression Eq. (3)] gives them a vital role in our model. Note, however, that, instead of explicitly extracting δ, τ, we use $A_{\nu\nu}$ in a scheme (see below) which dynamically *selects* those neural operators which can actually compute κ, either locally (Sect. 2.3.1) or bilocally (Sect. 2.3.2).

Joint selection of local ridge δ, τ and gauge frame: the S-field. The mechanism for finding the local 'ridge line' and its τ (Fig. 3) parallels the conventional scheme for finding local 'line orientation' (sensu neurobiology) in luminance images: at any μ, the (idealised) goal is to select the local maxima of $A_{\nu\nu}$ along the ν-dimension that are also global maxima across the useful τ-range (which covers the range of local slants β, see below). Almost any neural stage with lateral inhibition across these parameter ranges[12] does what is actually needed

(often called a 'winner-take-all' or 'iceberg' response): to reduce the $A_{\nu\nu}$ field to a field $S(\mu, \nu, \tau)$ which is positive only within a certain (μ, ν, τ) neighbourhood through which the nominal 'ridge-line' and its τ pass (see below and Fig. 4). There is no need to reduce the widths of this neighbourhood to less than roughly $\sigma_\epsilon, \sigma_\delta, \sigma_\tau$; in fact, trying to 'sharpen' S further risks losing noise robustness and/or network dynamical stability.

The S-field plays the role of selecting ('gating') the subset of neural κ-operators (Sect. 2.3) which have the locally appropriate (u, ν)-frame. Note that S remains a distributed code for the nominal 'best-fit' ridge location $(\hat{\mu}, \hat{\nu})$ and its local $\hat{\tau}$. This is precisely what we need to properly generalise Eq. (3): the formal local *values* of ν and ν_μ should be replaced by their local distributions (unnormalised). In any case, the nominal $\hat{\tau}$ hardly varies across any relevant 'strip' around the nominal ridge line, and our S-gated κ-computations are robust to such perturbations.

Figure 4 illustrates the geometry of S-fields: its joint encoding of ridge 'lines' and their local τ corresponds to (actually generalises) the important differential-geometric notion of 'lifting a graph into the slope bundle' (e.g. Burke 1985; Ben-Shahar et al. 2003). In our setting, the (blurred) subset of (μ, ν, τ)-space where S is large (it is mostly zero) indicates the subset of neural RFs that can measure the local second-order structure required for computing κ. Clearly, this can only work if (μ, ν, τ) can also smoothly parameterise each of the
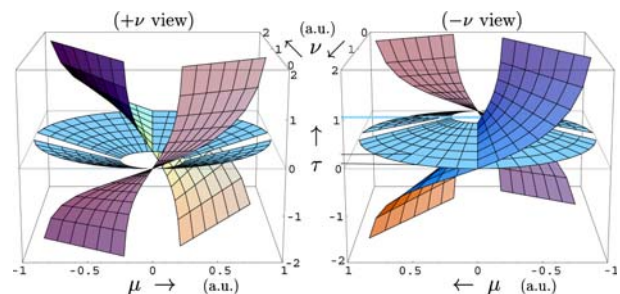
Fig. 4 The geometry of the S-field (here skeletonised for visualisation) which jointly extracts 'ridge lines' and their first-order contact parameter τ from the (μ, ν, τ)-tuned $A_{\nu\nu}$-field. To illustrate how such a scheme realises the differential-geometry notion of 'lifting curves into the τ-bundle', we use as data a set of circles through one point, all tangent to the μ-axis, but with different radii. The loci of large S-response then define the set of helicoidal *curves* shown. We also show these as projected onto the plane $\tau = 0$. To aid visualisation, we also draw in fictional 'spokes' at various fixed τ-values, as well as a fictional surface spanning the whole set of lifted *circles* (and their projections). Note also that we truncate the τ-range at ± 2, as identified before. This also shows up as the gaps in the projected set of *circles*. S-fields, such as rendered here by the 'lifted' curves, serve to select the neurons whose RF gauge frame (u, ν) fits the local ridge – only these neurons actually extract κ, locally or bilocally

---

[12] For example, the simple feedback scheme $\partial_t S = -S + [A_{\nu\nu} - \alpha \int d\beta \, G(\nu) \star S]^+$. Its steady-state response $S = [A_{\nu\nu} - \alpha \int d\beta \, G(\nu) \star S]^+$, has the required 'iceberg-response' properties for roughly $1 < \alpha < 10$. The reason for the integration measure $d\beta$ rather than $d\tau$ will become clear in Sect. 2.3.1.

eligible RF-families, but this is guaranteed by our construction (Sect. 2.2.2) of all Gaussian-derivative RF-families which respect the Galilean (μ, ν)-geometry. Picking any point in (μ, ν, τ) also singles out a local gauge frame (u, v) in which the local first-order structure vanishes (Fig. 3), thus setting the stage for measuring the local second-order structure. Using the finite-width S-field implies some local smoothing over sets of 'similar' gauge frames. This generalisation reduces the impact of τ-noise by averaging out its leading order effects, which are antisymmetric in the τ-error.

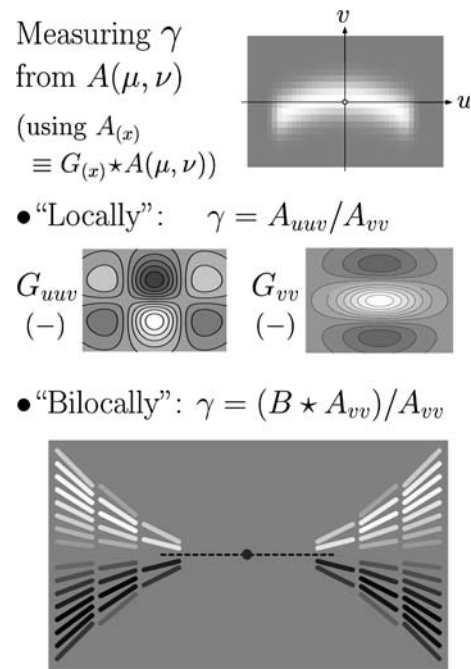### 2.3 Direct neural computation of object curvature κ

With $\nu_{\mu\mu}$ replaced by the second-order ridge-contact parameter γ, one could in principle compute $-a\kappa = W_1\gamma + W_0$ by first extracting γ and then correcting it (the view-dependent W_0, W_1 can be computed separately). However, the psychophysical data (Domini et al. 2001; Duke and Wilcox 2003) exclude this: adaptation is driven not by $\nu_{\mu\mu} \approx \gamma$, but by the perceived (possibly illusory) κ. In a two-stage computation, the γ-driven adaptation of the first stage would propagate through the second (κ) stage, which could complicate, but not undo, the effect. Hence, the corrections should be applied in the first stage that codes local second-order structure: we must *directly* extract κ.

Before designing the required RFs, we note that there exist two distinct schemes ('local' or 'bilocal') for extracting γ, analogous to the well-founded schemes (Koenderink and Richards 1988) for extracting monocular line curvature (Fig. 5). We modify these schemes to directly extract κ. We work out both schemes since they yield different predictions for neurophysiology. For psychophysics, the two will be hard to distinguish – unless perhaps by analogy with the line-curvature experiments of Wilson and Richards (1989).

2.3.1 Local scheme: view-modulated RFs acting on the A-field

To convert the known *local* scheme,[13] for extracting ridge-curvature γ into a scheme that extracts $-a\kappa = W_1\gamma + W_0$, without computing γ as a neural intermediate,

Fig. 5 Recalling the classic choice between local and bilocal schemes (Koenderink and Richards 1988) for extracting 'line curvature', such as our γ. A similar choice exists for extracting κ, but this requires dynamically modified RFs, derived in Sect. 2.3.1 and Sect. 2.3.2, respectively

we write $-a\kappa$ as

$$-a\kappa = \frac{W_1 G_{uuv}\star A + W_0 G_{vv}\star A}{G_{vv}\star A}$$
$$= \frac{-(W_1 G_{uuv} + W_0 G_{vv})\star A}{A_{vv}}. \qquad (8)$$

Note that the numerator expression on the right-hand side predicts a novel type of RF with *viewing-geometry-modulated structure* $(W_1 G_{uuv} + W_0 G_{vv})$. Such 'dynamic' RFs could well be realised neurophysiologically by the view-dependent gain modulation found by Trotter et al. (1992, 1996) and Trotter and Celebrini (1999).[14] Measuring the full structure of the predicted RFs under a variety of viewing conditions would be a particularly strong test of our model. The RF structures at fixed W_0, W_1 are all related to a single exemplar RF by $(\sigma_\epsilon, \sigma_\delta)$-dependent scaling and a (u, v) gauge transformation, just as the basic G_{vv}- and G_{uuv}-RFs from which they are built. The set of possible subfield patterns of these exemplars (Fig. 6) depends on only a single parameter, the 'mixing ratio' $\rho = (W_0/W_1)(\sigma_\epsilon^2/\sigma_\delta)$, which incorporates the

---

[13] To ease discussion and notation, we adopt henceforth the usual (Koenderink and Richards 1988) idealisation that the local gauges (u, v) are correctly fixed, unless stated otherwise. The errors can be shown to be negligible. In terms of our explicit mechanism (Sect. 2.2.3), this idealisation means that S selects ('gates') those neural κ-operators with (u, v)-frames that contact (order 0 and 1) the stimulus-evoked A-ridges. We do not actually write the S-multipliers acting on κ-operators.

[14] For example, writing the numerator as $-(G_{uuv}\star(W_1 A) + G_{vv}\star (W_0 A))$ shows how the reported type of gain modulation (of A-like responses) can do the required job. See Sect. 2.3.3 for further neural variants and details that are relevant to neurophysiological tests, but not to the computation as such.
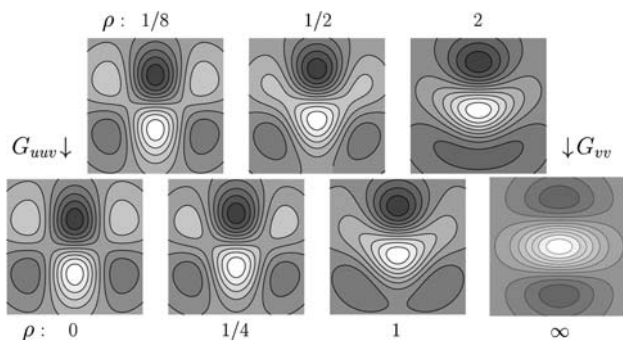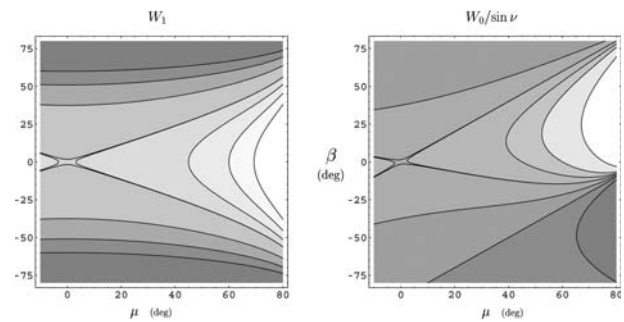
Fig. 6 RF patterns of local κ-detecting cells [numerator of Eq. (8)]): view-dependent mixtures of G_{uuv} and G_{vv}, parameterised by the 'mixing ratio' $\rho = (W_0/W_1)(\sigma_\epsilon^2/\sigma_\delta)$. Each RF is plotted here in its unit-scaled gauge-frame coordinates $(u/\sigma_\epsilon, v/\sigma_\delta)$. The rarely required RFs with $\rho < 0$ are obtained by jointly interchanging v (*up*) with $-v$ (*down*) as well as dark with light

Fig. 7 When parameterised by the cyclopean slant β, both W_1 and W_0 (normalised by $\sin \nu$) show only smooth dependence on μ and β. Strictly speaking, these plots show the far-field limits, but in β, μ-parameterised form, the exact result throughout the practically useful regime involves only minor, smoothly ν-dependent corrections. Symmetries: W_1 is symmetric in μ and β; W_0 is point-symmetric in $(0,0)$. Contour levels: $W_1 = \frac{1}{16}, \frac{1}{8}, \frac{1}{4}, \frac{1}{2}(\pm 10^{-3}), \frac{1}{2}\sqrt{2}, 1, \sqrt{2}$, and $W_0/\sin \nu = -1, 0, \frac{1}{2}, 1(\pm 10^{-3}), \sqrt{2}, 2, 4$

different scaling of the amplitudes of G_{uuv} and G_{vv} with the widths σ_ϵ and σ_δ. The mixed RF patterns measured under varying viewing conditions should thus map to a particular trajectory through the family of exemplar patterns sampled in Fig. 6, with the actual trajectory defined by the view dependence of ρ, which is determined via W_0 and W_1 by Eqs. (3) and (4). This suffices for testing the predicted RF patterns.

What is still missing, however, is a robust and plausible way of generating both W_1 and W_0 as required for the neural κ-computation,[15] in particular, the numerator of Eq. (8). This task is tightly linked to finding an appropriate neural sampling of the range of τ, which parameterises the gauge frames (u, v) in which each RF is defined (the widths $\sigma_\epsilon, \sigma_\delta$ are taken as constant until stated otherwise). At first glance, one might think that each neuron could be tuned to a fixed τ. However, as analysed in detail in Appendix A.3.2, this choice implies that the view dependence of the W_0 and W_1 signals required by all operators with $\tau \neq 0$ develops steep and sharp features in an otherwise smooth background dependence (Fig. 12 in Appendix A.3.2). This means that a fixed-τ population of neurons would suffer from sampling inefficiency and/or error sensitivity.

Geometric study of the source of these sharp features (Fig. 12 in Appendix A.3.2) suggests the (unique) way out of this problem: each neural operator involved in κ extraction should be assigned to a fixed slant β, rather than a disparity slope $\tau \approx \nu_\mu$. This choice is also supported by the analytical far-field expressions we found e.g. by $d \approx (\cos \mu / \cos \beta)^3$, the second form in Eq. (5).
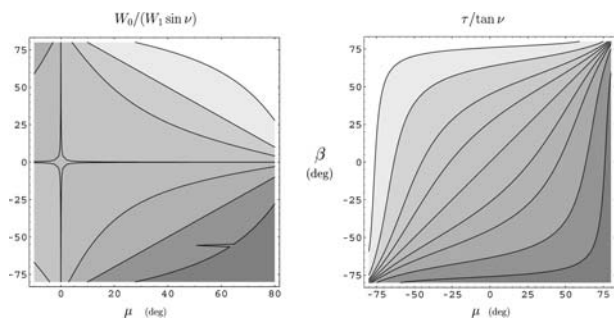
The expected advantages of fixed-β tuning are fully confirmed by the results shown in Fig. 7.

Indeed, W_1 and (distance-scaled) W_0 depend smoothly on β and μ over a range which easily covers all practical viewing conditions: even a neural dynamic range of about 10, say, allows a β and μ range of about ± 45, and $\pm 60°$ seems possible. Strictly speaking, these plots show the far-field limit ($\nu \to 0$) result, but another advantage of this parameterisation is that the result depends weakly and smoothly on ν over all practical viewing distances (breakdown only occurs near $\nu = \pi/2$).

Cell-fixed β tuning also implies that the gauge frames of all operators (including A_{vv} and S) are controlled by the now view-dependent $\tau = \tan \nu (\tan \beta - \tan \mu)$, the inverse of Eq. (2). Thus, it avoids drastic error sensitivity. Note also that the τ-global inhibition that transforms A_{vv} into S and the use of S in selecting the locally appropriate κ-responses are not affected – the latter only couples operators with the same (dynamic) (u, v).

The τ-driven gauge transformations can be implemented by the same gain modulation that performs the (W_0, W_1)-driven modulation of RF subfields. Transforming the (u, v)-frames boils down to a mere v-shift $v' = v - (\tau - \tau_0)u$, so the required gain-modulation signal at a point (u, v) in an RF is simply given by traversing (at 'speed' u) the fixed-u section through its exemplar RF (the effect for G_{vv} can be seen in Fig. 3).

Figure 8 summarises the two view-dependent factors that control all RF patterns under a fixed-β sampling scheme: $W_0/(W_1 \sin \nu)$ controls the RF subfield structure, and $\tau / \tan \nu$ controls each RF's dynamic gauge frame.

---

[15] In contradistinction to the one-parameter (ρ) dependence of the exemplar subfield *patterns*, the κ result also depends strongly on the overall RF amplitude. Thus, both W_1 and W_0 are needed (cf. Eq. (8)).

Fig. 8 The two parameters that dynamically define the RF-subfield patterns and gauge frames of κ-extracting cells (local or bilocal): the (scaled) RF pattern index $W_0/(W_1 \sin \nu)$ (*left*), and the (scaled) RF gauge parameter $\tau/\tan \nu$ (*right*) as functions of version μ and surface slant β. The $W_0/(W_1 \sin \nu)$ shown is strictly the far-field limit, but its error is negligible for any practical ν. The $\tau/\tan \nu$ *plot* is exact throughout. $W_0/(W_1 \sin \nu)$ is point-symmetric in $(0, 0)$. Contour levels: $W_0/(W_1 \sin \nu) = -4, 0, \sqrt{2}, 2(\pm 10^{-3}), 2\sqrt{2}, 4, 8$, and $\tau/\tan \nu = -4, -2, -1, -\frac{1}{2}, 0, \frac{1}{2}, 1, 2, 4$

2.3.2 Bilocal scheme: view-modulated RFs acting on the S-field

The basis of any 'bilocal' computation is that one can relate the order-$(n + 1)$ geometric structure at a point to the order-n geometric structure at pairs of (nearby) points. In our case, a local γ is related to nearby τ-pairs. Hence, we must first find the simplest viable scheme that extracts γ from properly selected pairs of τ-specific measurements, i.e. from the $A_{\nu\nu}$-field or its 'cleaned-up' version, the S-field. In fact, using the S-field as input takes care at once of the required selection of results computed in the proper local gauge frame (u, v) along the ridges.[16] Once we have a scheme that extracts γ, we can replace its static RFs by (W_1, W_0)-modulated RFs designed to directly extract κ.

In the Euclidean setting, the geometric demands on measuring the bitangent relation that defines bilocal curvature have been thoroughly analysed (Koenderink and Richards 1988), and similar ideas have been used in neurally inspired schemes for extracting regularised implicit curves or oriented texture fields (Ben-Shahar et al. 2003), but (unlike the local case) no robust neural scheme for explicit bilocal γ-extraction seems to be known. Specifically, we should avoid the proposed scheme where, in our case, γ would be found by subtracting locally extracted τ-values, since this introduces unbounded and unavoidable γ-errors as soon as one of the local patterns has an ill-defined τ-value. The way

out of such conundrums is simply to keep using distributed coding instead of explicit values. The design of our scheme goes as follows.[17]

If the centreline of an A-field ridge passes through some point (μ_1, ν_1) with local slope τ_1 and local curvature γ (like the example in Fig. 3), then the equation of the centreline's second-order contact element is the ('osculating') parabola $v = (\gamma/2)u^2$ in the local gauge frame (u, v). However, we have just concluded that a bilocal scheme must take its inputs from the S-field, which is carried by the three-parameter space (μ, ν, τ). Hence, we lift the osculating parabola $v = (\gamma/2)u^2$ to a curve in the (μ, ν, τ)-space (cf. Fig. 4, where this was done for sets of circles through a point), and we lift the (u, v) gauge frame to a $(u, v, \theta = \tau - \tau_1)$ frame. For the lifted osculating parabola, one simply has $\theta = \gamma u$, and the S-field in a neighbourhood of $(u, v, \theta) = (0, 0, 0)$ is approximately a blurred version of this lifted osculating parabola.

This setting indicates that RF-families suitable for our γ-scheme should be based on an exemplar with an RF pattern in resolution-scaled gauge coordinates $(u/\sigma_\epsilon, v/\sigma_\delta, \theta/\sigma_\tau)$. For example, it is reasonable to put the two formal sample points at $u = \pm \sigma_\epsilon$; at smaller u-distances, the dependence of the input field on γ drops quadratically, while at larger distances there could be undetected other structure between the sample points.[18] We need a roughly linear γ response over the range $|\gamma| < \sigma_\delta/\sigma_\epsilon^2$ allowed by the S-field resolutions.

The simplest exemplar RF pattern that fits our demands is (see top panels in Fig. 9)

$$B(u, v, \theta) = \frac{\sigma_\delta}{2\sigma_\epsilon^2} \Big\{ G(u - \sigma_\epsilon, v - \tfrac{1}{2}\sigma_\delta, \theta - \sigma_\tau)$$
$$+ G(u + \sigma_\epsilon, v - \tfrac{1}{2}\sigma_\delta, \theta + \sigma_\tau)$$
$$- G(u + \sigma_\epsilon, v + \tfrac{1}{2}\sigma_\delta, \theta - \sigma_\tau)$$
$$- G(u - \sigma_\epsilon, v + \tfrac{1}{2}\sigma_\delta, \theta + \sigma_\tau) \Big\}, \qquad (9)$$

where each $G(u, v, \theta)$ is a Gaussian with scales equal to (or a fraction of) those of its input S. Analogous to the local computation, $B \star S$ is also proportional to the amplitude scale of S, so γ is formally represented by the ratio of two RF responses (as already announced in Fig. 5). In the bilocal scheme, the required denominator RF is a simple G RF at the origin of the lifted gauge

---

[16] The fact that S-cells are actually β-tuned (by dynamic RFs) might seem to offer a direct route to κ, by comparing nearby S-pairs. However, this approach would reintroduce the ν_0-error sensitivity we avoid by our design, in which such errors only affect the usually minor *correction* of the robust γ to κ.

[17] The γ-extracting basic version of our scheme can also serve as bilocal scheme for line/edge curvature in monocular vision: simply replace our τ-tuned RFs with standard (Euclidean) 'angle'-tuned RFs.

[18] Note that such large-distance schemes can be very useful for creating smooth-shaped 'illusory surface' percepts from sparsely sampled data. We relegate such extensions of our scheme to later treatment.
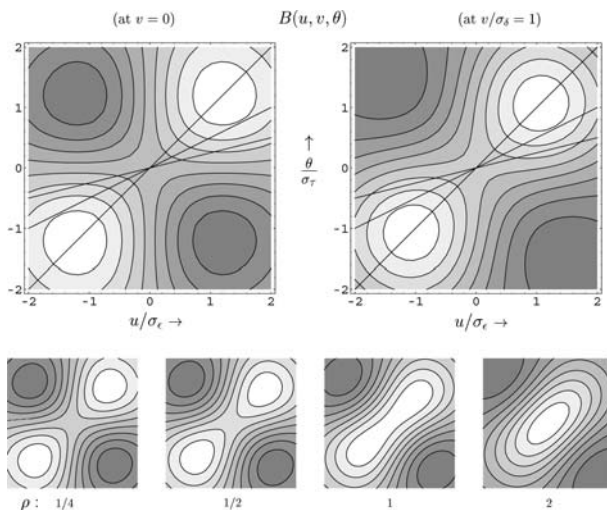
Fig. 9 RFs for bilocal γ- or κ-extraction from the S-field the 'cleaned-up' first-order contact code. *Top panels* Two sections through the bilocal RF $B(u/\sigma_\epsilon, v/\sigma_\delta, \theta/\sigma_\tau)$, which extracts γ from the S-field. As usual, we show all RFs in σ-scaled gauge-frame coordinates. An RF section at $-v$ is the dark-to-light and θ-sign inverted copy of a section at v. We have drawn in the projections on each section plane of a set of 'lifted' arcs with various γ. Such arcs only touch ($v = 0$) or intersect ($v \neq 0$) any constant-v plane, but the finite widths of the operators and their input field S itself cause the RF output to be quasi-linear in γ up to $|\gamma| \approx \sigma_\tau/\sigma_\epsilon = \sigma_\delta/\sigma_\epsilon^2$. *Bottom panels* For bilocal κ-extraction [Eq. (10)], the required RF for the numerator signal is $W_1 B + W_0 G$, shown here as some examples which cover the range of possible subfield patterns (at $v = 0$), indexed as before by the mixing ratio $\rho = (W_0/W_1)(\sigma_\epsilon^2/\sigma_\delta)$. RFs with negative ρ are obtained by jointly interchanging dark-light and up-down in the patterns shown

frame (u, v, θ). Thus we have formally $\gamma = B \star S/(G \star S)$. With this γ-scheme in mind, we have obtained a concrete reinterpretation of the lower right panel of Fig. 5: it visualises both the input stage (S: sloping bars, strongly shortened and thinned) and the B-weighted (grey-scale) summing over nearby S signals, in accordance with the family of curved local ridges parameterised by γ.

To convert this bilocal γ-scheme into one that directly measures κ, we rewrite it as

$$-a\kappa = \frac{W_1 B \star S + W_0 G \star S}{G \star S} = \frac{(W_1 B + W_0 G) \star S}{G \star S}. \quad (10)$$

The weights W_0, W_1, the RF mixing ratio ρ, and the dynamic τ-transformation of all RFs are identical to those in the local scheme. The bottom panel of Fig. 9 shows some samples from the range of $(W_1 B + W_0 G)$ RF patterns, in their scaled gauge frame.

2.3.3 Neural coding specifics: opponency, noise-robustness, etc.

Having determined all computationally essential aspects of our two schemes for κ-extraction, what remains is to

settle a few aspects of the neural coding that have little or no functional effect but that affect how the schemes we propose would manifest themselves in neurophysiological investigation.

We have already defined the neural A, A_{vv} and S signals such that they signal the relevant ('ridge') features while respecting the non-negativity of neural responses. What is still open is the neural encoding of κ, which we have so far written formally as a ratio of two neural signals in the local [Eq. (8)] as well as the bilocal [Eq. (10)] scheme.

Recall that κ, and thus also the numerators in either of our schemes, can have either sign. 'Opponent' pairs of neurons are a simple and suitable representation of bipolar signals, but there is no strict necessity to actually implement the formal division that would make κ a rate-coded value – an ordered pair of numerator and denominator signals is not only a perfectly adequate encoding of their formal ratio, but also has the major advantage that the denominator signal serves as a 'strength-of-evidence' measure for κ. At loci where a κ-value would be highly error prone, both signals are usually small, or they can be so scaled in a neurally straightforward way. The two-dimensionality of such a code avoids the many robustness problems which arise in value-coded computations because the latter falsely map a *'weak'* result into an *inaccurate* value.

If we choose the ordered-pair option, the numerator in either scheme would be encoded by an opponent neuron pair[19]

$$\text{Local}: \quad N_\pm = [\mp\{(W_1 G_{uuv} + W_0 G_{vv}) \star A\}]^+, \quad (11)$$

$$\text{Bilocal}: \quad N_\pm = [\pm\{(W_1 B + W_0 G) \star S\}]^+. \quad (12)$$

Note that the $N_\pm$-pair alone encodes κ confounded by e.g. the binocular luminance contrast or the 'sharpness' of the surface.

The other option (explicit κ-value) can then be regarded as a 'contrast-gain-controlled' version of the $N_\pm$-code. This can be done either in one step (using synaptic or membrane or spike-generation non-linearities to implement division or an equivalent, say logarithmic, representation) or in a distinct neural stage fed by $N_\pm$ and either A_{vv} (local scheme) or $G \star S$ (bilocal scheme). In either case, the simplest suitable neural implementation would be an opponent pair

---

[19] *Synaptic* gain modulation could directly implement the computation sequence indicated by the brackets written here, but the same result can be computed using the A-level gain modulation found by Trotter et al. (1992, 1996) and Trotter and Celebrini (1999), e.g. as $-(G_{uuv} \star (W_1 A) + G_{vv} \star (W_0 A))$. What matters [in view of the Domini et al. (2001) results] is that no intermediate neural stage carries a signal proportional to γ rather than κ. For example $G_{uuv} \star A$, or $B \star S$ is forbidden.

Local: $C_{\pm} = \left[\dfrac{\mp(W_1 G_{uuv} + W_0 G_{vv}) \star A}{\eta + A_{vv}}\right]^{+}$

or $C'_{\pm} = \dfrac{N_{\pm}}{\eta + A_{vv}}$, (13)

Bilocal: $C_{\pm} = \left[\dfrac{\pm(W_1 B + W_0 G) \star S}{\eta + G \star S}\right]^{+}$

or $C'_{\pm} = \dfrac{N_{\pm}}{\eta + G \star S}$. (14)

The new parameter η introduced here is necessary to prevent a 'divide-by-zero' type noise sensitivity. It should be of the same order of magnitude as the noise level of the main term in the denominator.

On the basis of present experimental data, we cannot decide between the various neural options we have worked out: the adaptation data (Sect. 1.2.3) (Domini et al. 2001; Duke and Wilcox 2003) allow either implicit coding by $(N_{\pm}, A_{vv})$-cell sets or explicit coding by $C_{\pm}$- or $C'_{\pm}$-cell pairs. Psychophysical testing is also expected to have trouble distinguishing between the local and bilocal schemes, since these are designed to be functionally equivalent. Neurophysiology, however, should be able to resolve these questions, in view of the neurally very different schemes that emerged from our two design variants. In any case, it is worth noting that all κ-representations, in both the local and the bilocal scheme, inherit from the RFs of their inputs a joint specificity for retinotopic disparity δ and its gradient δ_{ϵ} (actually, for surface slant β). There is already some psychophysical evidence for such joint tuning (Lee 1999; Aslin et al. 2004).

Moreover, all our mechanisms should show specificity for scale, of a novel type that should be neurophysiologically testable. As noted at several stages in the construction, the range of curvatures that can be extracted reliably and quasi-linearly (either locally or bilocally) scales with the widths of the constituent RFs as $|\kappa| < \kappa_{\star} \approx \sigma_{\delta}/\sigma_{\epsilon}^2$. Outside this range, the response saturates and eventually decreases with κ. [The scheme then also becomes more sensitive to errors in selecting the local gauge frame (u, v) and in the view-dependent modulation (mixing ratio ρ) of RF subfields.] The details differ somewhat per scheme, but we can focus on the common feature of the resulting representation: a large overall range of κ can be covered by copies of the same neural scheme with other RF-width combinations such that there is roughly at least one mechanism per doubling of $\kappa_{\star}$. This is illustrated in Fig. 10. Note the difference with respect to standard population codes which use an array of equal-width, shifted copies of an exemplar RF. Here, the RFs are all anti-symmetric in $\kappa = 0$ but different in their $\kappa_{\star}$-scale. Each RF alone allows a *pair* of κ-inter-
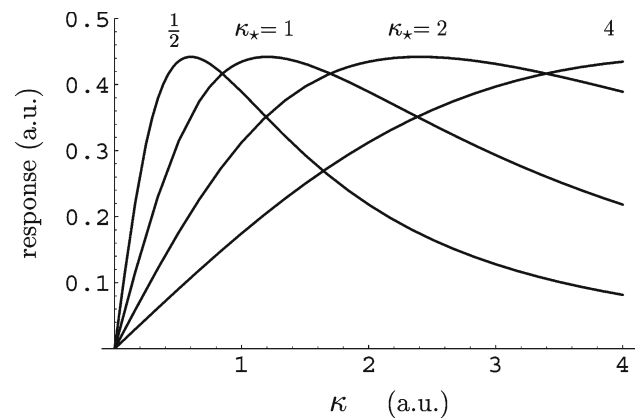
Fig. 10 Coding a large κ-range by a multiscale family of our shape extraction mechanisms. Each response is anti-symmetric in $\kappa = 0$, but differs in its quasi-linear range $\kappa_{\star}$, here by powers of 2

pretations, but the population code resolves this to a unique and well-defined κ. Note that this representation combines features of the classic 'opponent' as well as 'multichannel' coding schemes. This resolves the long-standing confusion (summarised in Howard and Rogers 2002) caused by trying to fit observations into this false dichotomy.

Finally, it is worth keeping in mind that the per-unit signal/noise demands on any neural version of our scheme are extremely mild, chiefly because of the vast number of units that contribute to producing the known surface-global (see Sect. 3) curvature Weber fraction. It suffices to make just a very rough order-of-magnitude estimate: the known Weber fraction (Rogers and Cagenello 1989) implies a signal/noise of at most 20 (i.e. roughly 4 bits of information) for *global* curvature-discrimination using very large parabolae that should stimulate at least 10^7 neurons. The Weber fraction scales roughly with the inverse of stimulus size. Thus, each unit needs to contribute only a tiny amount of curvature information, of the order of 10^{-6} bits. Further, if we assume that spatial integration is rougly linear and does not itself add much extra noise, then the per-unit signal/noise needs to be only 10^{-2}. Note that it is even no problem if much of the large (per-unit) 'noise' is actually systematic *error*, as long as those errors also are sufficiently uncorrelated to average out at the global level probed by the psychophysics.

3 General discussion

To develop a consistent, robust and potentially realistic model of perceptual shape from stereo, we started essentially 'from scratch': the vital first step was to

temporarily ignore all but a very focused selection of the vast psychophysical literature, rich as it is in unresolved conflicting interpretations, even when summarised as clearly as in the comprehensive review by Howard and Rogers (2002). Moreover, we found (e.g. Sect. 2.1.1) that all (metric or affine) depth- or slant-based approaches, and thereby almost all widely considered competing interpretations, are non-viable because they imply highly unrealistic sensitivity to vergence errors and/or violate crucial psychophysical or neurophysiological constraints. Having thus 'cleaned our slate', we let the whole path from theory to explicit neural model be dictated by geometric and computational analysis, within the constraints of just a few critical facts from psychophysics and neurophysiology.

Our construction may also be viewed as systematically exploring the implications of a sparse set of elementary (but interrelated) 'design choices'.

(i) *Direct extraction of local object curvature*, instead of (metric or affine) depth or slant. Extracting disparity curvature (Rogers and Cagenello 1989) is not enough: it must be corrected at once (Domini et al. 2001) to object curvature.[20] This involves compensating for several effects of the viewing geometry, specified by the local target vergence ν, version μ and surface slant β. The main error source (now only in the *correction* terms) is the formal fixation vergence ν_0, which actually combines several viewing distance signals including vertical disparity structure (Duke and Wilcox 2003).

(ii) *Using the appropriate ("Galilean") geometry of* (μ, ν)-*space*. Unlike Euclidean or affine geometry, this unites two basic constraints: the direction of (cyclopean) visual rays imposes the sign of local object curvature, and its value is available with good *metric* accuracy (Rogers and Cagenello 1989). The appropriate shape computations in this geometry then generate the novel families of modulated RFs and neural processes required for extracting object curvature from the (bi-)local structure of disparity data.

(iii) *Retinotopic population coding* by units that are jointly tuned for an expanding combination of local parameters: disparity [from the input $A(\mu, \nu)$ onwards], plus slant (from the gauge-field stages $A_{\nu\nu}, S$ onwards), plus object curvature (our out-

put stage). The finite spatial and parameter tuning width of these units automatically produces a realistic, robust form of surface interpolation and smoothing, capable of handling a sampled or 'transparent' depth structure.

Neural gain modulation by vergence and version signals then implements the dynamic view dependence of RF structures required for direct correction of disparity curvature to object curvature. There remain just two functionally equivalent but neurally different schemes for direct extraction of object curvatures: either locally or bilocally. Our neurally explicit designs for both variants make this choice decidable by neurophysiology.

At each stage of our design and its neural implementation, we have already specified at a detailed level the reasons for and consequences of each aspect of its structure. This also produced many detailed predictions, most notably the family of dynamically view-modulated RF subfields. Here, we put our approach and its implications in a wider perspective than was appropriate for specifying the computational and neural details. This includes making good on our implicit promise of offering an improved understanding of the many relevant psychophysics findings that we excluded from our design considerations. The most striking aspect of this is that our model actually predicts and explains the very data that were originally thought to plead *against* metric shape extraction.

3.1 Curvature-based metric shape versus existing proposals

A classic set of results which strongly favour curvature coding over depth coding is that, on the one hand, humans are essentially blind to global depth or disparity modulation (Erkelens and Collewijn 1985) and suffer from Cornsweet-type depth illusions (Anstis et al. 1978), while on the other hand they can discriminate curvature-in-depth with remarkably small Weber fractions, down to 5% (Rogers and Cagenello 1989; Johnston 1991). Note that such precision also argues strongly against *affine* depth or shape coding, which can only preserve the *sign* of curvatures.[21]

Moreover, computational robustness analysis (Sect. 2.1) has shown that any scheme that recovers *depth*-based shape from disparity entails unrealistically wild shape distortions caused by realistic errors on the (small) fixation vergence angle ν_0. These distortions would be far worse than linear depth scalings (affinities);

---

[20] As ultimate cause for direct *object* curvature extraction, we suggest the fact that κ, not γ, characterises the fixation-invariant correlation structure in binocular data generated by mostly rigid objects.

[21] See below for how our model also predicts that curvature *bias* is often larger than its Weber fraction.

they would even invert (or, at the very least, clip) the depth-profile of increasingly large portions of surfaces (from the distant side inwards) when one fuses stereo images that are slowly separated until some local target vergences ν become negative. Any slant-based scheme would show similarly unrealistic distortions.

Robustness to vergence errors only emerges when using the second-order disparity structure: as first noted by Rogers and Cagenello (1989), only the relation between disparity curvature and object curvature is free (to leading order) of the scaling by viewing distance that dominates the relation of metric depth (or slant) to disparity (or its gradient). Thus, curvature-based metric shape extraction has no need for the error-prone 'scaling' stage required in any depth- or slant-based metric shape computation.

A popular depth-based approach (e.g. von Helmholtz 1867; Gårding et al. 1995) proposed that the visual system may simply skip the depth-scaling operation and apply only a (more robust) 'disparity-correction' stage. This correction is superficially analogous to our correction of disparity *curvature* for the Vieth–Müller circle curvature and the effect of version, but the nature of the computation and its perceptual result is very different: disparity correction produces affine depth. Despite the existing evidence (see above) against affine depth coding, such models are widely thought to agree, at least loosely, with the perceptual errors reported in many (but not all) depth-interval or shape-constancy tests, in either purely binocular or multiple-cue settings (See Todd and Norman 2003 for the diversity of existing and new examples, which they interpret as excluding metric coding of 3D shape). Such interpretations also ignore that no affine-depth model actually explains these depth-task errors – for example, even the 'large' depth-to-width matching errors (up to about a factor of 1.5; e.g. Johnston 1991) are much *smaller* than predicted by affine coding, which only preserves the sign of the depth-to-width ratio.

3.2 Metric-shape extraction explains known depth-task errors

Our model puts all depth-task results in a very different light by predicting that their often 'contradictory' results in fact show various mixtures of three distinct types of error, each of which is a natural consequence of the direct extraction of metric curvature-based shape coding we propose. We first summarise the three error sources and their relevance to existing notions and experiments and then explain and discuss each item in detail.
(1) Direct extraction of metric shape does not involve depth (either metric or affine). Thus, psychophysical depth-task experiments do not actually probe our shape-extraction mechanisms directly. Moreover, any attempt to extract local depth or slant, either directly from disparity (Sect. 2.1.1) or indirectly from the robust shape code we propose, is fundamentally ill-constrained, in ways that turn out to fit much of the reported depth-task errors.
(2) Noise and bias in the vergence parameter ν_0 predicts global-scale errors in the correction of metric shape for e.g. the Vieth–Müller curvature. In most depth-task experiments, errors in metric curvature-based shape would be registered as errors in depth, in addition to the type-(1) depth errors.
(3) At large slant β, e.g. in strips along object contours, any extraction of object geometry must become error prone. This type of shape error predicts e.g. the relatively large errors registered in depth-to-width matching tasks.

It is impractical to address each of the many, often conflicting, interpretations and seemingly contradictory results that have emerged from a large variety of depth-task experiments (see for review, Howard and Rogers 2002; Todd and Norman 2003). More importantly, however, we can explain how our approach fundamentally changes the meaning and relevance of all such experiments. For instance, we can *invert* the widely accepted interpretation of large metric depth errors as evidence against metric and for affine coding of shape.

3.2.1 Even perfect metric shape often leaves depth and slant ill-defined

By construction, our mechanisms have the invariance properties which define the notion of shape (as appropriate to this setting). Indeed, we do not use or produce any depth- or slant-based object representation; the only use of (rough) estimates of viewing geometry and low-order disparity structure is to dynamically select the appropriate subset of curvature operators and modulate their RF subfields such that they directly extract metric object curvatures. In psychophysical terms, metric shape coding is inherently metameric for local depth and slant. Hence, the widely accepted interpretation of large perceptual errors in depth or slant as evidence against metric shape extraction is untenable. Indeed, the opposite interpretation applies, since even a metrically exact shape code actually *predicts* many of the observed errors, as we now explain.

Any ability to produce consistently stimulus-dependent scores in depth/slant tasks (often experienced as 'uncertain') indicates that the metamerisms left by pure shape extraction can be resolved to some extent. The potential sources of information for this process can be

classified as (combinations of) other cues, priors and the *non*-local depth/slant constraints implicit in a shape code. Here, we pay scant attention to the effects of priors (without denying their necessity in generating ecologically sensible responses to highly ambiguous stimuli); we focus on stimulus-driven means of resolving the metamerism. Their fundamental limitations yield the error predictions we are after.

Other cues. Using the local disparity or its gradient has been shown (Sect. 2.1.1) to produce wildly error-prone and distorted depth and poor-accuracy slant, except perhaps for very near viewing (up to a few times the interocular distance). Monocular cues are also unreliable sources of metric depth or slant. For example, motion parallax, shading, and texture carry their own affine depth ambiguity (or worse); accommodation cannot be accurate at most relevant distances; and using perspective depends on prior knowledge of the object. A large body of psychophysics [reviewed in Howard and Rogers (2002) and Todd and Norman (2003)] has confirmed these fundamental limitations. In summary, additional cues may weakly constrain the full local depth/slant metamerism of shape coding, but they generally leave the result metrically ill-defined and/or biased, independently of the potentially accurate metric shape code.

Nonlocal constraints. The potential use of depth and slant constraints implicit in any given shape code warrants far more theoretical and experimental study than we can spend on it in this paper, but we can sketch the core aspects that are of immediate relevance. To start, consider viewing a (projectively fixed) 'cut-out' of a sphere or cylinder (or plane) such that the object contour is not seen.[22] The complete shape code (i.e. all the object curvatures, as generated by extending our model to the full 3D problem) is constant across the visible surface patch and is thus unaffected by any change in the object distance, slant or lateral shift (at fixed tilt for the cylinder), as long as this does not move the object contour into the aperture. This turns the original *local* depth and slant metamerism of shape coding into a surface-*global* metamerism for these shapes. A much wider set of shapes allows similar metamerisms, albeit without the lateral shifts, or only if one allows (realistically small) shape-code tolerances.[23] In broad

summary, the prediction is that scaling the assumed or estimated mean distance of a patch (with small mean slant) induces a roughly quadratic scaling of its perceived depth range. We interpret the classic 'cardboard-cut-out effect' (Howard and Rogers 2002) as an extreme manifestation of this: When fusing at close distance a pair of stereo photographs of a large scene, each object appears to be nearly flat (without losing depth discontinuities). At such small distances, the usually weak accommodation and vergence cues should be strong enough to yield extremely underestimated object distances. Given well-determined local curvatures, the perceived depth range of each object is then predicted to be reduced much more strongly than the perceived lateral size, leading to a 'cardboard-cut-out' appearance (and the concurrent 'micropsia').

Most of the formal depth-/slant-task experiments under discussion used one or more of the simple smooth shapes which we found to have particularly strong distance, depth and slant ambiguity, globally within each object and independently across depth discontinuities. As such, these experiments cleanly probe the ambiguity-resolving ability of other cues or priors, *not* the accuracy of true shape coding. How limited these abilities are is revealed by the profound depth blindness (Erkelens and Collewijn 1985) and limited slant accuracy (Gillam et al. 1984; van Ee and Erkelens 1996) for planar stimuli, and we now predict that similar effects manifest themselves in tests using spherical, cylindrical or 'depth-grating' surface patches – where they have generally been misinterpreted as failures of metric shape coding. The specific predictions for each experiment vary, but they share a common predicted pattern: the full depth and slant ambiguities of pure shape coding will be constrained only weakly (and jointly) by the low-resolution slant representation that is apparently available (Gillam et al. 1984; van Ee and Erkelens 1996). Our mechanisms already use a low-resolution slant encoding, the *S*-field, to select operators that extract the properly slant-/depth-invariant shape code; we suggest that the same *S*-field be used also for the very different purpose of (weakly) constraining the surface interpretation of the field of shape codes.

Integrability errors. There is one more central aspect of the shape-code metamerism, and it predicts an essentially richer class of depth and slant errors than predicted by the above analysis. This concerns the question whether a given shape, with its possibly complicated

---

[22] The no-contour demand can often be replaced by weaker but less easily phrased demands; our aim here is just to give the simplest relevant examples. Note also that our Sect. 3.2.3 below deals with the different type of errors arising in the vicinity of visible object contours.

[23] For instance, all smooth 'depth-grating' stimuli allow a large range of joint distance and depth-profile changes – bounded only by the fact that large increases in distance eventually require some local slants β to reach $\pm\pi/2$, contradicting the absence of a contour

there. Moreover, *any* smooth and uniformly small-slant shape is nearly ambiguous for view-distance scaling with proportional lateral scaling and quadratic depth-profile scaling, until the implied local slants are no longer negligibly small.

pattern of contours or surface creases, actually offers sufficiently strong non-local constraints to robustly reduce the vast joint ambiguity of all *local* metamerisms to a few-parameter ambiguity per object, as discussed above. Formally, it takes one spatial integration of local shape to yield slant and a second integration to yield depth. For any path of integration, this introduces two (for depth, or one for slant or depth interval) free parameters, the two 'constants of integration'. For sufficiently smooth shape fields and contours, the freedom of all path-dependent parameters is heavily restricted by the (patch-*global*) demand of integrability: in essence, the integration along any loop on a smooth surface should not produce a jump in slant or position. If this constraint could be implemented without error, it would reduce all local ambiguities to the few-parameter metamerisms per object, as explained above. However, even smooth surfaces can have contours that strongly restrict the size and shape of loops around which the integrability demand can actually offer an effective constraint. Moreover, any realistic implementation of the integrability constraints produces errors which must increase as the paths get longer, and especially if it crosses sharp curvature variations (say, surface creases). Prime examples of the latter situation occur in Craik–Cornsweet stimuli, which indeed produce particularly strong bias on formally local depth scores, but one should expect similarly context-dependent depth biases to occur much more generally. It is also worth noting that *sets* of depth scores across objects can violate any affine relation to the object depths. Indeed, when integration errors become dominant, there is nothing to prevent sets of depth scores from violating any globally consistent depth transformation, including the global depth *order*.

All of the above predictions can be seen to explain major patterns in the variety of depth/slant errors reviewed in Todd and Norman (2003). Integrability errors in particular explain the recently found violations of any global depth transformation (Bingham et al. 2004) which firmly contradict all previous theories of shape from stereo.

3.2.2 From viewing distance bias, via curvature bias, to depth bias

The other two types of depth error we predict stem from true shape errors, independently of the metamerism-type errors, and both involve a common source which lies *outside* our model but is vital to its operation: the fixation vergence parameter v_0. Measuring a small difference between the two eye positions, each of which is noisy and biased (Steinman et al. 1982), v_0 cannot avoid being inaccurate under practical conditions,

where viewing distance is at least an order of magnitude larger than interocular distance.

The first relevant effect is that a v_0-bias induces a global bias in our viewing-distance-dependent curvature correction (W_0) for the curvature of the Vieth–Müller circle through the fixation point. A similar (but roughly half-amplitude) bias is predicted for vertical curvature when extending our neural computations to the full 3D case, which involves compensating for the shape of Vieth–Müller tori. Viewing-distance-dependent curvature bias is a well-known contribution to failures of depth-constancy. The classic example is the objective curvature of the 'subjective frontoparallel' (von Helmholtz 1867), but it is often disregarded in interpreting similar lacks of constancy as evidence against metric shape coding (Todd and Norman 2003). Biased Vieth–Müller correction due to direct or vertical-disparity-induced v_0-bias is revealed most clearly by the curvature-adaptation experiments (Domini et al. 2001; Duke and Wilcox 2003) which helped to pin down our choice of shape representation. It should also contribute a bias term to the classic 'cardboard-cutout effect' (Howard and Rogers 2002) with originally convex objects, in addition to the symmetric depth-range compression we predicted earlier; the two terms could be separated by comparing concave and convex objects.

3.2.3 Shape failure near contours implies depth-width matching errors

To understand the final type of error, imagine first that one could find the exact curvature field across the cyclopean projection of a generic smooth object, including its contour. Integrating the curvature field with the slant boundary condition $\beta = \pm \pi/2$ along the contour then generally pins down some metric object diameter, which – in combination with its visual diameter – would pin down the object distance, and thereby the complete metric situation. However, this ideal is clearly beyond the reach of any real implementation: any finite-resolution operator must fail near the κ-discontinuity at $\beta \pm \pi/2$. More subtly, the v_0-errors induce errors in the required slant-dependent correction of disparity curvature $v_{\mu\mu}$ to object curvature κ (correction for μ is much less error prone). This error sensitivity also gets large where $|\beta| \approx \pi/2$. The net effect is not only to block access to object-diameter and viewing distance, but also to the depth range covered by the near-contour strip (at least one RF-width wide in projection), where any computation must break down. In other words, it is entirely expected that there are large errors in attempting to match the depth range of most objects to their width. Todd and Norman (2003) contains many examples, read as evidence against

the existence of metric shape coding]. Note also that the essential loss of the near-contour κ data as a global shape constraint effectively returns us to the (partial) metamerisms explained in the previous two sections, except for a generally weak additional constraint on the object distance. An informative example (Johnston 1991) is where large errors (bias up to 50%) were found for adjusting the shape of elliptical cylinders, presented at various distances, to perceived circularity, whereas low Weber fractions (7%) were found for curvature discrimination in the same setting. Nearby (far) shapes were seen with enlarged (reduced) depth, consistent with our model (as well as others) under the reasonable assumption that subjects underestimate the range of viewing-distance variations used.

3.3 Other psychophysical tests of curvature-based processing

Some detection or discrimination experiments have also been interpreted as evidence against early extraction of second-order disparity (or object) structure. In particular, Lunn and Morgan (1997) varied the wavelength and waveform of depth gratings, and Petrov and Glennerster (2004) varied the scale of dot triplets, and found that the dependence of discrimination performance on these manipulations does not correspond well with how they affect the formal disparity curvature in the stimuli but correlates better with their relative disparity (or other options, which differed between the experiments). This was interpreted as excluding curvature-based processing – a conclusion that actually only applies to models that use rate-coded units with equal RF size. Any computationally and neurally viable model must employ a considerable range of RF sizes. In our schemes, operators with specific spatial (σ_ϵ) and disparity (σ_δ) scales are part of a multiresolution set of κ-tuned responses that combines the classic features of 'opponent' and 'channel' coding (see end of Sect. 2.3.3). In its logically simplest form, this construction yields scale-invariant Weber-law behaviour, roughly the very pattern of results which was mistaken for evidence against curvature-based processing. Moderate deviations from this ideal are expected because of a finite range of RF sizes and scale-dependent per-unit signal-to-noise ratios.

Adaptation experiments have provided much more crucial information for model construction than other psychophysical data, because the former can probe directly the structure of the data representation and processing, while the latter are often so confounded by other factors (see above) that interpreting such data *requires* rather than constrains a mechanistically explicit model. As a positive case in point, just a few informative

adaptation experiments (Sect. 1.2.3) (Domini et al. 2001; Duke and Wilcox 2003) essentially narrow down the choice of shape representation to retinotopic curvature with immediate ν_0-controlled correction, but without any coding of depth. As we saw earlier in this discussion, this basic choice of representation already puts a huge collection of hitherto confusing depth-task data in a completely new light. Our model also fits adaptation data which were originally interpreted in terms of a slant-based model (Lee 1999): these experiments show clearly that binocular shape processing uses disparity-tuned operators, and they are compatible with a slant-based model, but do not require it – they are equally compatible with our types of curvature-based units, in which slant tuning does not serve to extract slant as such but to represent the gauge field S that selects the proper curvature operators and determines the immediate correction of disparity curvature to object curvature.

Our model suggests future adaptation studies specifically aimed at the fact that our curvature-extracting RFs have an 'orientation' τ in (μ, ν)-space. This predicts shape adaptation to be selective for 'slant', even though the shape percept is coded by curvatures, with the overall slant weakly and inaccurately constrained by the first-order disparity and the viewing parameters. The degree of selectivity of curvature adaptation for slant β is defined via Eq. (2) by the RF aspect-ratio σ_μ/σ_ν, given the viewing geometry μ_0, ν_0. RFs with various aspect ratios will contribute to perceived after-effects, so one cannot hope to extract the tuning curve of any single RF, but the occurrence of at least some slant selectivity in curvature adaptation for near viewing, and its broadening at larger viewing distance, should be measurable.

3.4 Neurophysiological relevance

Neurophysiology is probably the most informative means of testing our model. Our input data representation $A(\mu, \nu)$ is already well supported by existing recordings from area V1: binocular cells are retinotopic and usually disparity-tuned (with the classic notions of 'near' and 'far' cells now identified as merely having large σ_ν). The finding that such cells preferentially sample *horizontal* disparity (Cumming 2002) was thought to be 'unexpected' but is precisely what one expects for curvature-based (but not any other) shape extraction: as Koenderink and van Doorn (1991) proved (for motion, with stereo as a special case), the second-order structure of the full (i.e. 2D-vector) disparity field is purely epipolar.

Recent recordings (Nienborg et al. 2004) found that the RFs of V1 binocular cells have large aspect ratios ($\sigma_\mu \gg \sigma_\nu$), with the long axis roughly along constant

disparity lines. This supports our assumptions for the input stage $A(\mu, \nu)$. Note also that it is perfectly compatible with our approach if cells at the $A(\mu, \nu)$ level have some inhibitory surround around an (elongated) Gaussian main lobe – this only affects how these RFs effectively interpolate and smooth the raw disparity signals to produce the potentially transparent 'surface' representations on which our curvature-extraction schemes operate.

The next stage, requiring specific RFs that extract local curvature from the initial disparity encoding, has not yet been found experimentally, or at least not fully characterised. In constructing our neural implementations (Sect. 2.3), we have fully specified the two predicted families of RF structures and how the RF of any given cell should be dynamically modulated. This rich set of predictions invites detailed testing. Already we can note some promising similarities with existing neurophysiological results obtained by Trotter et al. (1992, 1996). They found that viewing distance (linked to our ν_0) gain-modulates many binocular V2 cells, usually in a monotonic way, consistent with the modulation of RF-subfield pattern which our model uses to correct for Vieth–Müller curvature. The neural data can in fact be interpreted as showing particular crosssections through such an RF, but the full structure remains to be measured. In further experiments (Trotter and Celebrini 1999), neural modulation was found to be driven also by version μ_0, often with a 'one-humped' dependence. This is again tantalisingly close to what our model predicts in correcting for version and slant. Indeed, some of the data already suggested dynamic 'rotation' of RFs, roughly as our prediction of the dynamic τ-transformation of each RF's gauge frame (u, v), parameterised by the constant β to which each curvature-extracting unit is tuned. Similar but less specific neurophysiological results have been reported from other areas (Rosenbluth and Allman 2002).

More neurophysiology experiments along the lines of those reported by Trotter et al. appear to be a very promising route to testing the many specific neural predictions of our model and to exploring how it is embedded in a wider network which must include not only the machinery that supplies ν_0 or its equivalent but also some stages beyond shape extraction. Indeed, the shape coding in area IT (Janssen et al. 1999, 2000, 2001) is interpretable as a stage beyond our curvature-based shape coding. The computational features added in this next stage would then be retinal position invariance, as well as competition between co-local shape units responding to highly different κ ranges. Recent findings of slant-tuned cells (Nguyenkim and DeAngelis 2003; Hinkle and Connor 2002) were interpreted only in terms of an assumed functionality of slant per se, but they could just as well be part of our curvature-extraction stage, or some next stage in the processing of shape, since our approach predicts all such cells to be also tuned for slant.

3.5 Completing and extending the model

We have limited this paper to the problem of extracting the curvatures of intersections of objects with epipolar planes. This subproblem actually already presents one with all the computationally new and essential aspects of the full 3D task, but in their simplest form possible. Generalising the model to compute the full 3D structure does not require essentially new notions or techniques (but will make most of the formulae look much more complicated), nor will its implementation require essentially different neural machinery – one simply needs to add essentially the same neural units but rotated in at least two additional directions of visual 'tilt'. The curvature corrections for all tilted curvature operators then contain extra terms due to the vertical curvature of the equal-vergence tori and due to essentially the classic Meusnier correction (Koenderink 1990) for the local tilt/slant combination.

A natural next stage of extending the model would be to integrate it with modules that extract shape specifically near contours, where the present computation becomes error prone. This is also the level where *occlusion* first becomes a relevant consideration, albeit in a highly non-standard manner: small-scale 'roughness' of many realistic opaque surfaces dissolves the visually relevant 'occluding contour' into an effectively semi-transparent strip – a situation that dovetails nicely with our present approach, which is free of opaqueness assumptions. Beyond that strip lies the monocular region which can support 'da Vinci stereopsis' (Pianta and Gillam 2003, and references therein). Note that generalisation of our model to fit this form of stereopsis would require extending the Galilean geometry we chose here to capture the *cyclopean* structure of (μ, ν)-space. Changing the formal geometry would change the RF families (since these are generated by its rigid motions, as in Sect. 2.2.2), but only in the so far nearly irrelevant regime $|\nu_\mu| > 2$. Most of the formal geometry we used in our present model was developed mathematically in the 1940s (e.g. Strubecker 1942) forms the basis for our application), and it has occasionally been used in computer-vision (Pottmann and Opitz 1994; Koenderink and van Doorn 2002) and in visual science (Koenderink 2003), but for different purposes and in different guises – always because it correctly respects the special role of one dimension (say, of visual rays) with respect to that of others (say, the two dimensions of the visual

field). The fact that it allows metric shape (i.e. curvature fields) instead of just projective or affine structure, despite the special status of the 'visual ray' dimension, makes it the proper geometry for cyclopean binocular vision. This contrasts with the earlier visual applications of this geometry, which concerned *monocular* information that leaves metric shape undefined. In any case, our model appears to be the first use of this geometry for designing a neurally explicit model that extracts visual metric 3D shape. It should be useful also as a basis for integrating this information with that from locally even more ambiguous cues than disparity.

Appendices

A Shape from disparity, vergence and version

A.1 Notation, and coordinate transformations (zeroth order)

We restrict analysis to the geometry of curves in an arbitrary plane of regard (Fig. 1), generalised to guarantee 3D genericity (Sect. 2.1), with $(x, y > 0)$ as Cartesian coordinates (right, ahead), and with eyes at $(x, y) = (\pm a, 0)$. The cyclopean visual coordinates are (μ, ν), the version and (full) vergence, with the (generalised) fixation point at (μ_0, ν_0). Note that $y > 0$ implies $2|\mu| + \nu < \pi$. The Vieth–Müller circle, of radius R, has its centre at $(0, y_C)$. Viewing distance D is defined from the 'cyclopean eye' at $(0, y_C - R)$.

These quantities are mutually connected by the transformations

$$x = a\frac{\sin 2\mu}{\sin \nu}; \quad y = a\frac{\cos 2\mu + \cos \nu}{\sin \nu};$$
$$R = \frac{a}{\sin \nu}; \quad y_C = a \cot \nu; \quad D = 2R \cos \mu \tag{15}$$

$$2\mu = \arctan\left(\frac{x+a}{y}\right) + \arctan\left(\frac{x-a}{y}\right)$$
$$= \arctan\left(\frac{2xy}{y^2 - x^2 + a^2}\right);$$
$$\nu = \arctan\left(\frac{x+a}{y}\right) - \arctan\left(\frac{x-a}{y}\right)$$
$$= \arctan\left(\frac{2ay}{y^2 + x^2 - a^2}\right). \tag{16}$$

Classic models (zeroth order) try to recover 'depth' (either D or x) from these relations, despite their obvious

sensitivity divergence in the realistic small-ν regime. See Sect. 2.1.1 for a detailed explanation of how all zeroth-order methods must produce severe shape distortions and instability when faced with realistic errors in the fixation vergence ν_0.

A.2 First-order structure, and Galilean (μ, ν) geometry

Local first-order shape is often expressed in terms of the HC slant α or slope $y_x \equiv -\tan(\alpha)$, but this choice leads to expressions (Koenderink 1992 and below) with singularities that have no visual significance. The only choice that avoids such problems is to compute the slant β relative to a plane normal to the cyclopean visual direction μ. Actually, the simplest expression is obtained for $\tan \beta$, as follows.

At distance D along a cyclopean visual ray (direction μ), let T measure distance orthogonal to it (such that $T_\mu > 0$). Thus, we have

$$D_\mu = -2R \sin \mu, D_\nu = 2R\frac{\cos \mu}{\tan \nu},$$
$$T_\mu = 2R \cos \mu , \ T_\nu = 0, \tag{17}$$

which allows us to find

$$\tan \beta = -\frac{dD}{dT} = -\frac{D_\mu d\mu + D_\nu d\nu}{T_\mu d\mu} = \tan \mu + \frac{\nu_\mu}{\tan \nu}. \tag{18}$$

This simple result plays two vital roles in designing our model, first in showing why slant-based models also do not have perceptually relevant robustness (Sect. 2.1.1), and second, in determining how the (μ, ν) parameters control the dynamic transformations of RFs that extract a curvature-based shape code (Sect. 2.3.1 and beyond).

The more familiar HC slope $y_x \equiv -\tan \alpha$, with $\alpha = \mu + \beta$, can now be expressed as $-y_x = (\tan \mu + \tan \beta)/(1 - \tan \mu \tan \beta)$, which illustrates why $\tan \beta$ is not only geometrically but also analytically a more natural quantity than $\tan \alpha$: it avoids the divergence at $\tan \mu = 1/\tan \beta$, which signals the cases where the surface tangent is parallel to the x-axis – an event without any perceptually special status, unless $\mu = 0$.

Returning to the use of local slopes $\tan \beta$, one may note that this, rather than slant β, is the first-order measure that is natural to the *visual* geometry, in either monocular or cyclopean binocular vision. Indeed, the slope discontinuity at $|\beta| = \pi/2$ corresponds to a visual discontinuity: even for a semi-transparent or sparse-sampled local surface, the (cyclopean) view switches from one side of the surface to the other. For opaque objects, monocular occlusion intervenes when $|\nu_\mu| \geq 2$, at slopes which readily follow from Eq. (2). The same critical slopes emerge in our κ-analysis [Eq. (3), or Sect. A.3

below], but without any role for occlusion (which is simply non-existent in our derivation).

The visually defined sign (convex/concave) of surface curvature κ must also invert at $|\beta| = \pi/2$. This elementary fact implies that the visually natural geometry of our half-plane differs structurally (i.e. not just because of the choice of the μ, ν coordinates) from that of the familiar Euclidean x, y-plane in which rotations over any angle β are a natural 'motion'. The fact that Euclidean rotation of a rigid object eventually violates the continuity of its visual shape representation (Sect. 2.1.2) shows the fundamental incompatibility between the two geometries. Nevertheless, extracting *metric* shape is possible (and actually implemented by our model), except when $|\beta| \approx \pi/2$, i.e. near κ-discontinuities. This situation (almost everywhere metrically defined shape) is fundamentally beyond what can be handled by the widely assumed 'retreat' from metric to affine, or even weaker-structured surface representations.

Note that the visually appropriate intrinsic structure of our half-plane singles out the ν-direction, i.e. the cyclopean visual rays, as special. The deep consequences of this simple fact seem not to have been widely appreciated. The most important aspect for our application is that any kind of 'rigid motion' in the proper (μ, ν)-geometry must leave all lines parallel to the ν-axis invariant, as illustrated by the notion of 'rotation' (τ-motion $(u', v') = (u, v + \tau u)$) used extensively throughout the design of the model (from Sect. 2.2.2 onwards). This geometry may seem unusual, but it is actually well-known to any high-school graduate: it is the (1+1)-D Galilean space-time geometry of Newtonian physics. In mathematics, the formal structure that fits the required (μ, ν)-structure is called 'isotropic geometry' (Strubecker 1942; Sachs 1987), a particularly unfortunate historical accident, since all other sciences call such structures 'anisotropic'. In view of our intended readership, we avoid using the mathematical name throughout this paper. Introductions and summaries of the geometry relevant to the relatively basic aspects that we use in this paper (Sects. 2.1.2 and especially 2.2.2, with sufficient local explanations) are available (Pottmann and Opitz 1994; Koenderink and van Doorn 2002), but one should note that these papers use the geometry differently, as befits the very different applications they aim at.

A.3 Object curvature κ

No correct expression relating the object curvature κ to the disparity curvature $\nu_{\mu\mu}$ appears to have been published,[24] except for the special case ($\mu = 0, \beta = 0$) (Domini et al. 2001, but see below), so we compute the exact result from the basic transformations Eqs. (15) and (16) connecting the Cartesian and binocular coordinates.

Since we only have to deal with the intersection curves of surfaces with a plane of regard, our intended result can be found from the classical expression for the curvature κ along a planar curve $F(t) = (x(t), y(t))$ with (general) parameter t:

$$\kappa(t) = \frac{x_t \, y_{tt} - x_{tt} \, y_t}{\left[x_t^2 + y_t^2\right]^{3/2}}. \tag{19}$$

In particular, we want to express κ at a point (say, $t = 0$) of the $F(t)$-curve in terms of all the local parameters of the curve $\phi(t) = (\mu(t), \nu(t))$, which is the cyclopean binocular image of $F(t)$. The special status of cyclopean visual rays excludes considering points where $\phi(t)$ is tangent to a constant-μ line. Combining these conditions with the smoothness of the binocular mapping in the domain of interest, we may choose a local general $\phi(t)$-model as $\mu(t) = \mu + t$, $\nu(t) = \nu + \nu_\mu t + \frac{1}{2}\nu_{\mu\mu}t^2$, where $\mu, \nu, \nu_\mu,$ and $\nu_{\mu\mu}$ are now fixed parameters (since only local shape near $t = 0$ matters here). Transforming $\phi(t)$ back to $F(t)$ via Eq. (15), we can find κ by evaluating Eq. (19). To obtain the result in a compact and comprehensible form, it is useful to do the same for some intermediate steps (computer algebra systems tend to produce very large, unclear and impractical representations of the result). The numerator of Eq. (19) then yields

$$-\frac{a^2}{\sin^3 \nu} \Big[2\nu_{\mu\mu}(\cos 2\mu + \cos \nu) \\ -(\nu_\mu^2 - 4)(\nu_\mu \sin 2\mu + 2 \sin \nu) \Big], \tag{20}$$

while the term $x_t^2 + y_t^2 = (x_\mu + \nu_\mu x_\nu)^2 + (y_\mu + \nu_\mu y_\nu)^2$ in the denominator can be reduced to

$$\frac{4a^2}{\sin^2 \nu} \times \Bigg\{ 1 + \frac{\nu_\mu \sin 2\mu}{\sin \nu} + \Big(\frac{\nu_\mu}{2 \sin \nu}\Big)^2 \\ \Big[(\cos 2\mu + \cos \nu)^2 + \sin^2 2\mu\Big] \Bigg\}. \tag{21}$$

Already at this point, it is worth noting the special case $\nu_\mu = 0$, which corresponds to tangency of the object with a Vieth–Müller circle – a situation which may be

---

[24] The original proposal (Rogers and Cagenello 1989) used an approximation which is only correct up to leading order in viewing distance and restricted itself to $\mu = 0$ and $\beta = 0$. The only published general result (Koenderink 1992), given without derivation, is unfortunately incorrect, as can be seen from its divergences and its failure to reproduce simple special cases.

rare in nature but is often used in psychophysical stimuli. This makes the denominator a constant and leads to the much simpler exact result

$$-a\kappa = \frac{\nu_{\mu\mu}}{4}(\cos 2\mu + \cos \nu) + \sin \nu \quad ; \quad \nu_\mu = 0. \quad (22)$$

Further specialisation to $\mu = 0$ simplifies considerably the only previously solved special case (Domini et al. 2001).[25] This allows one to quantify the first type of problem of the Rogers and Cagenello (1989) approximation (as already sketched in Fig. 1) that $\kappa \propto \nu_{\mu\mu}$: to keep the curvature error below their measured 5% Weber fraction, $|a\kappa|$ must remain 20 times larger than the term $\sin \nu = a/R$ above. Thus, when viewing, say, a sphere, the viewing distance $D \approx 2R$ must be at least 20 times the sphere diameter. Equivalently, the visual angle subtended by the sphere may not exceed about 3°. This also implies that the dynamic RF substructure modulation (mixing ratio ρ) for Vieth–Müller correction used in our neural κ-operators is indeed functionally important for all RFs with total width $\approx 2\sigma_\epsilon$ exceeding 3°. Judging by the widths measured (Nienborg et al. 2004) already at the A-field level (area V1), this should indeed apply to the neural stages (V2 or higher) where we expect our model to be realised.

Returning to the general case, one can put the exact result into a relatively compact form which also brings out the structure we exploit in designing the neural model (Sect. 2.1.2 and beyond)

$$-a\kappa$$
$$= \frac{\frac{1}{4}\nu_{\mu\mu}(\cos 2\mu + \cos \nu) + \left(\sin \nu + \frac{1}{2}\nu_\mu \sin 2\mu\right)\left(1 - \frac{1}{4}\nu_\mu^2\right)}{\left\{1 + \frac{\nu_\mu \sin 2\mu}{\sin \nu} + \left(\frac{\nu_\mu}{2\sin \nu}\right)^2 \left[(\cos 2\mu + \cos \nu)^2 + \sin^2 2\mu\right]\right\}^{3/2}}.$$
$$(24)$$

A.3.1 'Far-field' regime: $\nu \to 0$ at finite $\nu_\mu/\nu \equiv \xi$

Most practical viewing conditions are in the 'far' field, where approximating the results to first order in ν is accurate enough. Note, however, that simply taking a $\nu \to 0$ limit at any finite ν_μ leads to divergent or meaningless results; all realistic situations require that $\nu_\mu/\nu \equiv \xi$ remain bounded as one takes $\nu \to 0$. All other cases correspond to an asymptotically 'grazing' view of a (receding) surface, singular cases which we have already

---

[25] In the context of conic sections defined by a geometric parameter H which equals $-\nu_{\mu\mu}/2$ in our notation, the result given was (using our notation)

$$\kappa = \left(\frac{1}{2a}\right)\frac{4\tan \nu/2}{1 + \tan^2 \nu/2}\left(\frac{H}{2\tan \nu/2} - 1\right). \quad (23)$$

This equals our simple result $a\kappa = \frac{H}{2}(1 + \cos \nu) - \sin \nu$.

found not to allow curvature extraction from disparity. Under this bounded-ξ proviso, we obtain

$$-a\kappa = \frac{\frac{1}{2}\nu_{\mu\mu}\cos^2 \mu + \nu\left(1 + \frac{1}{2}\xi \sin 2\mu\right)}{\left\{1 + \xi \sin 2\mu + (\xi \cos \mu)^2\right\}^{3/2}} + \mathcal{O}(\nu^2). \quad (25)$$

To reveal the large-slant behaviour discussed in the main text (Sect. 2.1.2), specifically Eq. (5), it is useful to rewrite the denominator (d) in this result. Using $1 = \sin^2 \mu + \cos^2 \mu$ and $\sin 2\mu = 2\sin \mu \cos \mu$ yields

$$d = \{\cos^2 \mu + (\sin \mu + \xi \cos \mu)^2\}^{3/2}$$
$$= \cos^3 \mu\{1 + (\xi + \tan \mu)^2\}^{3/2}. \quad (26)$$

Noting from Eq. (18) that in the far field $\xi + \tan \mu = \tan \beta + \mathcal{O}(\nu^2)$, and using the general equality $1 + \tan^2 z = 1/\cos^2 z$, we find the simple result $d = (\cos \mu/\cos \beta)^3 + \mathcal{O}(\nu^2)$.

If we further restrict the parameters to small slants and versions, the curvature expression reduces to

$$-a\kappa = \frac{\nu_{\mu\mu}}{2}\left(1 + \frac{\mu^2}{2} - \frac{3\beta^2}{2}\right)$$
$$+ \nu\left(1 + \mu\beta - \mu^2\right) + \mathcal{O}(\nu^2, \mu^2, \beta^2), \quad (27)$$

which can serve to illustrate another error incurred under the original proposal of approximating $\nu_{\mu\mu} \propto \kappa$: even for far viewing, where the additive (Vieth–Müller) term is small, and at small μ, there is only a small range of slants (roughly $|\beta| < 12°$) where the multiplicative error is as small as the 5% Weber fraction on curvature.

A.3.2 Computationally robust reparameterisation of the κ-computation

As noted already in Sect. 2.1.2, Eq. (24) is linear in $\nu_{\mu\mu}$, with coefficients depending on all other parameters. Thus, we write it as

$$-a\kappa = (b\nu_{\mu\mu} + c)/d = W_1\nu_{\mu\mu} + W_0. \quad (28)$$

This leaves us with the task of generating the weights W_1 and W_0. Their formal dependence on μ, ν and ν_μ [cf. Eq. (24)] may appear ill-suited to neural computation. To prove the contrary, we analyse this dependence more closely. This yields a formally equivalent parameterisation which smoothly and efficiently covers the perceptually relevant range of viewing situations. It also provides insight as well as neural predictions.

To get a first impression, it is useful to look at the simple special case $\nu_\mu = 0$, which corresponds to a surface patch which is locally tangent to a Vieth–Müller circle. This sets $d = 1$ and $b = \sin \nu$, which leaves us with very smooth weights $W_1 = (\cos 2\mu + \cos \nu)/4$ and $W_0 = \sin \nu$, as illustrated in the two top panels of Fig. 11.
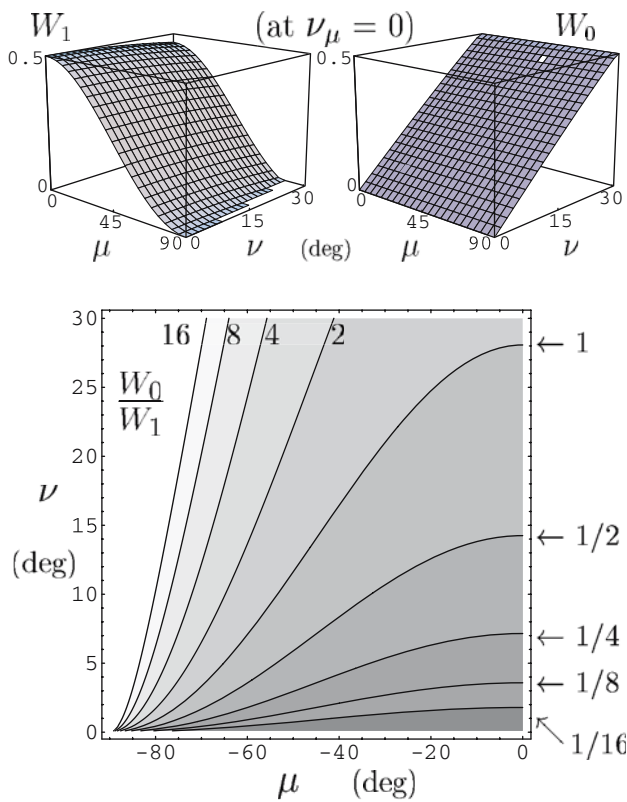
Fig. 11 Version and vergence dependence of 'weights' W_1 and W_0 (*top left/right*) which should gain-modulate the receptive field contributions for computing κ, and, (*bottom, contourgraph*) the resulting 'relative correction' W_0/W_1 which (partly) controls the resulting RF subfield pattern. The μ-symmetry is even for each graph; the ν-symmetry is even for W_1 and odd for the others. Note: this introductory example is atypically smooth because $\nu_\mu = 0$ sets $d = 1$

This case provides the simplest possible example of how our κ-computation adds the required viewing-distance correction ($W_0 = \sin \nu$) to the original proposal (Rogers and Cagenello 1989) of using γ as a rough approximation of κ. The 'relative correction' W_0/W_1 (bottom panel of Fig. 11) is the view-dependent factor in the mixing ratio ρ which indexes the dynamic RF subfield structure (Fig. 6) in our local scheme. Note that the W_0/W_1 ratio here is also smoothly dependent on μ and ν, except for unrealistically eccentric μ.

However, the case $\nu_\mu = 0$ is atypically 'well behaved'. An example of the complications that arise for any fixed non-zero ν_μ can be seen in Fig. 12. W_1 (left panel) develops two sharp features not seen so far: first, a small-ν 'boundary layer' now connects the moderate-W_1 regime to the limit $W_1 \to 0$ at $\nu = 0$ (the effective ν-thickness of this layer scales as ν_μ, so the case $\nu_\mu = 0$ is actually *singular* in the far-field limit). Second, as μ increases (with sign opposite to that of ν_μ), an increasingly sharp and high 'spur' occurs on W_1. All these effects can be fully understood by formal analysis of Eq. (3), but geometric

insight into their cause can be gained from the right-hand panel of Fig. 12, which shows the (μ, ν)-image of (x, y)-circles (30-cm diameter) centred at eccentricities $\mu = 0$, 22.5, and 45°, and depths D of 1, 2, 4, and 8 m (for $a = 3.2$ cm). We also show the image of two constant-D curves at $D = 1$ and 2 m and a line segment with $\nu_\mu = -1/30$ to indicate the fixed RF slope used in the W_1 plot.

The crucial aspect is how the images of the (smaller than D) circles deform as D increases: their diameter *along* the constant-D curve image through their centre scales roughly as $1/D$, while their *transverse* diameter scales roughly as $1/D^2$, turning them into near-ellipses with an aspect ratio (long/short axis) which grows roughly as fast as their long axis shrinks. This implies (see figure) that the local $\nu_{\mu\mu} \approx \gamma$ along much of the length of the ellipses grows and becomes steeply μ-dependent (which corresponds to where W_1 becomes small and/or steep), except where their tangent is roughly parallel to a local constant-D curve, i.e. where the local (cyclopean) slant β is moderate.

The steep and narrow features in an otherwise smooth μ, ν dependence for any (non-zero) fixed τ imply that using RFs with a fixed τ would suffer from sampling inefficiency or error sensitivity. However, the geometric insight offered by Fig. 12 suggests that β should be a more 'natural' parameter than τ for the RFs used for κ-extraction. This is also supported analytically by the far-field analysis (Sect. A.3.1 above), especially by the relations

$$b = \frac{1}{2} \cos^2 \mu + \mathcal{O}(\nu^2), \quad d = (\cos \mu / \cos \beta)^3 + \mathcal{O}(\nu^2) .$$

Some further algebra yields $c/\sin \nu = \cos^2 \mu + \frac{1}{2} \sin 2\mu$ $\tan \beta + \mathcal{O}(\nu^2)$. These far-field approximations turn out to require only weak and smooth corrections for all practical ν (breakdown occurs only near $\nu \approx \pi/2$). Thus, not only $W_1 = b/d$ and $W_0/\sin \nu = c/(d \sin \nu)$, shown in the main text in Fig. 7, but also the distance-scaled RF pattern index $W_0/(W_1 \sin \nu) = c/(b \sin \nu)$, shown in the main text in Fig. 8, left panel, are smooth functions of the visually meaningful parameters μ and β. The same goes for the scaled first-order contact parameter $\tau/\tan \nu$ (shown in Fig. 8, right panel) which controls the dynamic gauge frames (u, v) of all neural operators.

B Adaptation-transfer experiment details

The aims of the experiment and a summary of its design, results and interpretation have been given in Sect. 1.2.3, and an abstract was presented at ECVP (Noest et al. 2003). Here we complete the description.
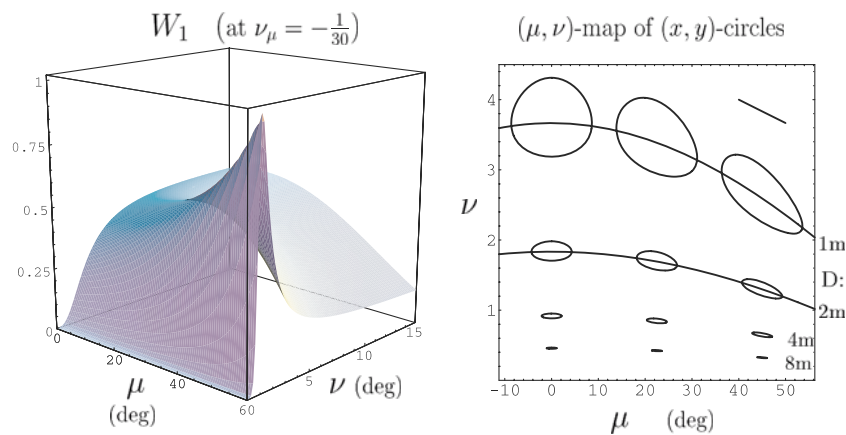
Fig. 12 Example showing why a *fixed* τ-tuning is unsuitable for κ-computing neurons. *Left panel*: the weight W_1 for any fixed non-zero $\nu_\mu \approx \tau$ develops sharp features: a small-ν 'boundary layer' and a 'spur' at increasing μ. *Right panel*: the geometric source of the sharp W_1 (and W_0) features becomes apparent by plotting the

(μ, ν)-images of (x,y)-circles at various distances ($D = 1, 2, 4, 8\,\text{m}$) and two constant-D curves ($D = 1, 2\,\text{m}$). See text for detailed interpretation. In short: robust, visually relevant shape information occurs only where local slant β is moderate
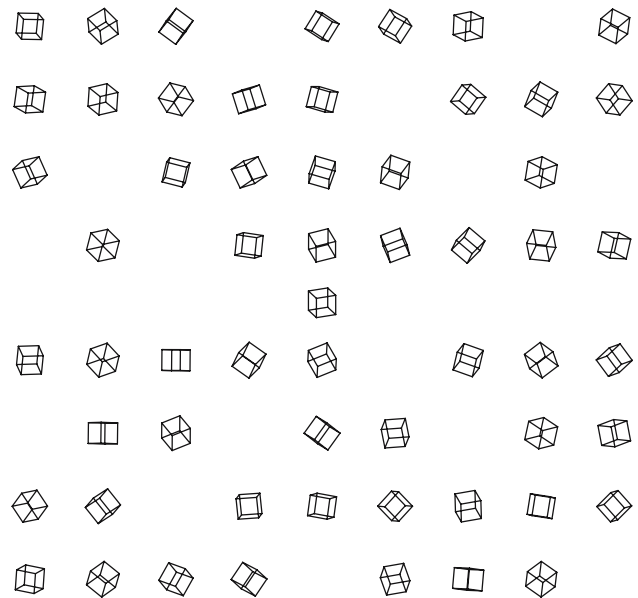
B.1 Methods and stimuli

B.1.1 Measuring adaptation strengths

In all cases, a basic adaptation measurement consisted of a staircase procedure which adjusts the objective curvature K of a test stimulus until it nulls the perceived aftereffect of a fixed-curvature adaptation stimulus. Before each staircase, adaptation was built up for $100\,\text{s}$ (well beyond the period required for effective 'saturation' of the after-effect), and each test stimulus ($2\,\text{s}$) was followed by a 10-s adaptation 'top-up'. To quantify adaptation per se (as separate from curvature bias), we used paired consecutive staircase runs in which the adaptation stimulus was a convex/concave version of the otherwise fixed adaptation shape (either a vertical cylinder or sphere, see below). The order of using the convex/concave shape within each pair of runs was random.

B.1.2 Adaptation and test stimuli

The formal surface shapes of the convex/concave pair of adaptation stimuli were cylindrical with vertical axis (in four of seven experiments, see below) or spherical (in remaining cases), with formal curvature $\kappa = \pm\kappa_0 = \pm\frac{4}{3}\text{m}^{-1}$. Independently of the azimuthal position of the stimulus centre (at $\mu = \pm 15°$), the cyclopean slant was kept at $\beta = 0$ and the viewing distance at $1.5\,\text{m}$.[26] Head movements were restricted by a bite board.

Fig. 13 Example stimulus, subtending 30 by 30°, centred at $\mu = \pm 15°$ version, and rendered as a 10^6 pixel (at 75 Hz) *red/green anaglyph*, projected on a frontoparallel screen at 1.5 m (equal to the simulated viewing distance at the stimulus centre). Elements: 58 ± 4 wire-frame cubes with 1° (≈ 2.7 cm) sides, randomly 3D-oriented and placed on any of the $2 \times 9 \times 4 = 72$ potential positions in a two-block square grid, as shown. Fixation element: one wire-frame cube, either moving horizontally (see text) or static and centred, within the otherwise empty central row

The formal shape of each stimulus patch (adaptation and test) was actually rendered as a moderately sparse, semi-random array of wire-frame cubes centred on the formal surfaces, as illustrated in Fig. 13, with details of the layout and rendering given in the figure

---

[26] Thus, the *concave* stimulus has $|\nu_{\mu\mu}|$-values very much smaller than that of the convex stimulus.

legend. Apart from the display, the room was dark and offered only a very dim view of black drapery.

Given this rendering of the stimuli, any RF with σ_ϵ up to about 2° can only sample local disparity (or object) structure that bears no consistent relation to anything that could cause the shape-adaptation effects we quantify in our experiments. Only RFs with σ_ϵ between about 3 and 7° sample the stimulus we use at a scale where the adaptation effects we measure can arise at all. At these width scales, there is also a substantial contribution to the RF-subfield structure from the dynamic Vieth–Müller correction (mixing ratio ρ).

One of our aims is to quantify the extent to which shape is coded in a retinotopic or a HC frame. Contributions from several stages might well produce superimposed effects, but we can separate these by comparing a 2×2 subset of conditions generated by independently presenting the adaptation and the test stimulus in the Left or Right visual hemispheres (see below).

B.1.3 Static versus moving fixation

Finding the nature of the local quantity that codes shape requires another doubling of the number of conditions, differing in how subjects fixate the adaptation stimulus (the test-stimulus fixation is always fixed in its centre). One condition is static fixation, on a stimulus element (see below) in the centre of the simulated surface. The other condition consists in *tracking* the fixation element as it moves on a smooth 3D path which intersects the simulated shape, spanning 80% of its width and several times its depth range. The motion along this path is a Lissajous trajectory with a 2:1 frequency ratio of its sinusoidal motion components in depth and width respectively. One full Lissajous figure is traversed during each adaptation 'top-up' phase. This moving fixation condition serves to 'scan' any retinal-frame mechanisms across a wide range of local values of disparity and depth relative to the fixation; the same goes for the gradient (respectively slant) for the convex stimuli.

The situation is illustrated in Fig. 14, showing a.o. the two adaptation stimuli (convex/concave) with the Lissajous path used in the 'moving fixation' condition and a few test stimuli, all mapped into binocular coordinates (version μ and vergence ν). For example, the large disparity variation which occurs, say, at the fovea during the 'moving fixation' condition appears here as the varying (and sign-inverting) ν-distance between a point on the Lissajous path and its constant-μ projection on either of the two adaptation stimuli. Likewise, one can see the large local ν_μ-variations encountered during the lateral component of the fixation scanning. However, the local
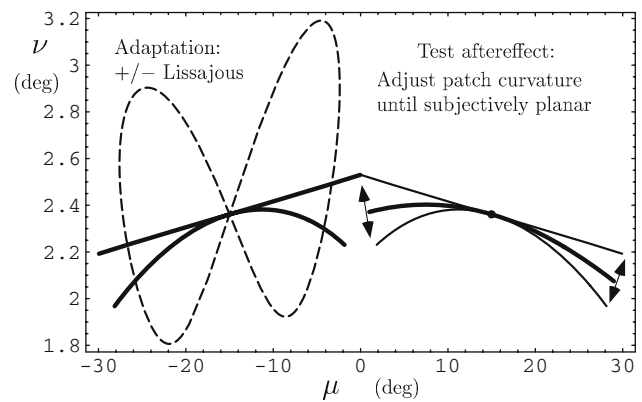
Fig. 14 Formal stimulus surfaces (*left*: adaptation; *right*: test) and the Lissajous fixation path (*dashed*), each projected onto the plane of regard and mapped into the binocular (version,vergence)-space (μ, ν). The example shown corresponds to the conditions testing L→R adaptation transfer. Note that the concave (radius 0.75 m) adaptation stimulus maps to a (very nearly) straight line in (μ, ν), while the convex stimulus maps to a line with negative $\nu_{\mu\mu}$

second-order structure (either $\nu_{\mu\mu}$ or κ) is almost constant during the Lissajous-path fixation scanning.

As intended, the effect of the Lissajous-figure fixation conditions is to 'smear' the local adapting signals in non-curvature-based units so widely across the set of all possible units that it would very strongly reduce the *difference* in the adaptation state across that set as caused by the opposite-shape pairs used. Indeed, our Lissajous tracking smears disparity values over a range (measured as $\pm\sigma$ of the distribution) of about eight times the convex–concave mean difference; for the local disparity gradient, the smearing is about five times its difference. The disparity curvature $\nu_{\mu\mu}$ is much less smeared (about one quarter of its concave–convex difference), while κ is, of course, not affected at all.

B.2 Subjects, results and analysis

We collected 7 full data sets (4 with cylinder and 3 with sphere stimuli) using the 16 conditions listed in Sect. 1.2.3. The 7 datasets were produced by 5 stereo-competent subjects with normal or corrected to normal vision; all except AJN (author) were paid volunteers, studying physics or biology, but naive about the meaning of the experiments. To obtain each dataset, a subject made at least 4 repeat runs through each of the 16 experimental conditions, taking about 12–15 h to complete a set, divided up into 2-h sessions.

To address our questions about the nature of the shape code, we regroup the data from the 16 adaptation conditions into 4 groups:

Group (index i)	Fixation (adapting)	Headcentric left/right (adapt → test)
1	Patch-centre	(L→L) & (R→R)
2	Lissajous	(L→L) & (R→R)
3	Patch-centre	(L→R) & (R→L)
4	Lissajous	(L→R) & (R→L)

Thus, the first (last) two groups contain the data from conditions without (with) Left/Right 'transfer' between hemispheres.

Within each group, the data occur in pairs of raw adaptation measures K_+, K_-, measured after adapting to the convex or concave version of any of the stimuli. As mentioned earlier (Sect. 1.2.3), our interest focuses on the differences $\Delta K = K_+ - K_-$, but for completeness, we also compute the sums, which measure something like curvature 'bias'. It makes sense to average these measures over all pairs within each group $i = 1..4$, and express the result relative to the adaptation curvature κ_0. Thus, we reduce the data to

adaptation 'strength' $S_i = \frac{\langle K_+ - K_- \rangle}{2\kappa_0}$;
'bias' $B_i = \frac{\langle K_+ + K_- \rangle}{2\kappa_0}$, (29)

where $\langle \cdot \rangle$ means the within-group averaging described above.

As already illustrated in Fig. 2, the answers to our questions are strikingly clear from the pattern in scatterplots of S_3 versus S_1, and S_4 versus S_2, with their respective SEM values represented as symbol widths. For our purposes (choosing the coding format for our model), this semi-quantitative evaluation is perfectly adequate. For completeness, we add here the results of extracting some numerical patterns of interest from the data.

Specifically, we linearly transform the four S_i (or B_i) into four new components:

$M = [Z_1, Z_2, Z_3, Z_4]$, $D = [Z_1, Z_3] - [Z_2, Z_4]$,
$H = [Z_1, Z_2] - [Z_3, Z_4]$, $X = [Z_1, Z_4] - [Z_2, Z_3]$, (30)

where Z_i is either S_i or B_i, and $[\cdot]$ signifies optimally weighted average over the list of indexed arguments in the brackets. Note that M is the global mean, D senses coding based on disparity (or its gradient) rather than on the object curvature, H does the same for HC rather than retinotopic coding, and X senses bilinear interaction between D and H. We also compute the SEM of each component.

From the data measured with cylinder stimuli, we then find the component values given in Table 2 (with Z_1 for reference). Likewise, from the data measured using spherical stimuli, we find the component values given in Table 3.

Focusing on the adaptation strengths, which are relevant to our questions, these results clearly support the conclusions we already read from the scatterplots in Fig. 2: all components D, H, X which might signal the presence of HC and/or non-curvature-based coding are well below the global mean adaptation strengths M. For four of the seven datasets, D, H, X are at least ten times smaller than M, indicating that retinotopic curvature-based coding is by far the dominant contribution. For the other three datasets, the D, H, X terms are *statistically* significant (roughly three to four times their SEM), but still two to seven times smaller than the global mean adaptation strength M, and even smaller relative to the more model-relevant reference measure S_1. What matters to our model is capturing the dominant contribution, which is clearly consistent with retinotopic curvature-based representation. Moreover, small D, H, X terms are actually not unexpected even in a purely retinotopic curvature-based system, since they arise as side effects of the neural implementation, as explained below.

B.3 Interpretation details (main aspects in Sect. 1.2.3)

The dominant pattern in how adaptation depends on $\mu = \pm 15°$ transfer and/or moving fixation clearly hints at retinotopic curvature-based coding (Sect. 1.2.3), but one might wonder whether the sometimes (in three of seven datasets) detectable, small to moderate deviations from the idealised pattern hint at the existence of other shape representations operating in parallel with the mechanism that we modelled. However, realistic side effects of the jointly disparity- and slant-tuned neural implementation of our model can produce these types of deviations. This is discussed in the first two subsections.

Finally, we discuss the considerable curvature bias we found and explain why it is of no relevance to our model itself but can be explained by the demands it (and any model) places on other visual processing, involved in estimating viewing distance.

B.3.1 Side effect of opposite-μ adaptation transfer

The small but sometimes detectable H-term in our data might reveal a weak contribution from a head-centrically coded stage. Such an interpretation has indeed been given to similar small gaze-dependent effects found very recently (Nishida et al. 2003) for motion, tilt and size adaptation. However, small effects of this type are expected to arise as side effects of realistic neural implementation, certainly in our case, and possibly also in the other cases just mentioned.

Table 2 Values of adaptation strength components [defined in Eq. (30)], measured using cylinder-patch stimuli

Subject	AJN		LvdW		MLS		NK	
	Strength	Bias	Strength	Bias	Strength	Bias	Strength	Bias
$Z_1(\%)$	24 ± 3	18 ± 1	21 ± 2	30 ± 7	34 ± 2	12 ± 3	18 ± 2	5 ± 5
$M(\%)$	23 ± 1	19 ± 1	16 ± 1	26 ± 3	34 ± 1	9 ± 2	12 ± 1	13 ± 2
$D(\%)$	2 ± 3	1 ± 2	5 ± 1	4 ± 6	3 ± 2	2 ± 3	6 ± 1	-4 ± 4
$H(\%)$	-1 ± 2	-1 ± 2	5 ± 1	-4 ± 5	-1 ± 3	2 ± 3	5 ± 1	-1 ± 3
$X(\%)$	1 ± 2	-1 ± 2	-4 ± 1	6 ± 5	-1 ± 3	3 ± 3	-2 ± 1	-4 ± 3

Table 3 Values of adaptation strength components [defined in Eq. (30)], measured using spherical-patch stimuli

Subject	AC		AJN		MLS	
	Strength	Bias	Strength	Bias	Strength	Bias
$Z_1(\%)$	14 ± 2	16 ± 2	25 ± 1	14 ± 1	26 ± 2	7 ± 3
$M(\%)$	13 ± 1	18 ± 2	22 ± 1	15 ± 1	22 ± 1	7 ± 2
$D(\%)$	1 ± 2	-5 ± 3	5 ± 1	0 ± 1	1 ± 1	3 ± 4
$H(\%)$	1 ± 2	-7 ± 5	3 ± 2	-5 ± 2	2 ± 2	-2 ± 4
$X(\%)$	-2 ± 2	-2 ± 4	3 ± 2	1 ± 2	1 ± 1	0 ± 3

In our case, consider first the 'same' versus 'opposite hemisphere' conditions, both with static fixation: as we showed in Fig. 14, the adaptation and test stimulus mapped into (μ, ν)-space differ in their global-scale disparity slope ν_μ by about 0.2. The same is true in the neurally relevant retinal (ϵ, δ)-coordinates. This global ν_μ-mismatch also implies a δ-mismatch which grows with ϵ across the two types of stimuli. Both aspects can contribute to a loss in overlap between the set of units which is driven by the adaptation stimulus and the set which gets 'read out' by the test stimulus. For the largest RFs encoding the shape (these units must be centred near the stimulus centre) loss in overlap will occur only when the global slope difference exceeds the τ-resolution of the RFs. This limits the loss of overlap to RFs with aspect ratios $\sigma_\epsilon / \sigma_\delta$ larger than about 5, but recent data (Nienborg et al. 2004) suggest that this is not unrealistic. For smaller RFs (but still larger than a few times the 1° size of the stimulus elements) which sample the more eccentric parts of the stimulus, the δ-mismatch can also cause loss in overlap of the adapted and tested set of units, if σ_δ is smaller than about 0.15°. Quantifying these effects is unfeasible because it depends on the unknown distribution of units across the doubly-multiscale version of (μ, ν, τ)-space, and because adaptation may well spread out in (μ, ν, τ)-space beyond the set of cells which are directly driven by the adaptation stimulus, but one can certainly conclude that even our purely retinotopic, curvature-based model is likely to show a small loss in adaptation measure S_i under the 'opposite hemisphere' condition.

B.3.2 Side-effect of moving fixation during adaptation

Next, consider the slight loss in ΔK when going from the 'static' to the 'moving' fixation condition (say, while adapting and testing in the same mean visual location). Again one can see from Fig. 14 that there is now a considerable difference between the adapted and the tested set of cells (at the end of Sect. B.1, we gave measures of the ratio of the RMS ranges of the disparity and its slope which are 'scanned' across by our Lissajous-path-fixation procedure). We note the following properties of the set of adapted cells and their expected consequences for our ΔK measures:

(1) The disparity ranges of the adapted sets of cells for the convex and concave stimuli overlap so strongly that it would reduce ΔK to near-zero for any shape code based on disparity values (or their slopes, which also become smeared). The fact that our data show only a slight loss in ΔK thus strongly favours shape coding by curvature, which is not smeared by Lissajous fixation – the local curvature-driven adaptation in these sets of cells is retained no matter whether they overlap in their disparity tuning (also, it is irrelevant here that we actually implement the curvature sensors as opponent cells). (2) The set of tested cells lies entirely within the set of adapted cells. Thus, the same total exposure to the adaptation stimulus is now spread across a larger set of cells. In fact, for a reasonable example with $\sigma_\delta = 0.25°$, the adapted set is roughly three times larger, or more if adaptation spreads beyond the high-activity set. Per cell, one then expects a certain reduction in the 'strength' of

its adaptation. A three-fold reduction would be possible if the 'static' adaptation level were well below saturation, but we chose the adaptation period (100 s) to be well beyond the time needed to reach saturation (roughly estimated in pilot runs with static fixation). Thus, the neural adaptation state should not drop much below saturation even when the cells effectively 'see' the adaptation stimulus for about one third of the time. Perceptually, the result should be only a minor loss in ΔK, as is indeed observed. Thus, the model again predicts at least qualitatively a deviation from the ideal which, at first sight, might have been read as evidence for the presence of a different neural process than the one we used in building the model.

B.3.3 Curvature bias caused by external v_0-parameter estimate

The existence of a perceived curvature bias is perhaps the most classic viewing-distance-dependent imperfection of shape-from-stereo – in essence, the old 'horopter problem' (von Helmholtz 1867). Intentionally, we focused on the adaptation measures ΔK (and the S_i) designed to null out any bias, and our model design also ignored it, for reasons we now explain. First, note that a considerable bias indeed occurs in our data (see the B_i-derived measures in B.2), as well as in other data (Domini et al. 2001; Duke and Wilcox 2003) we used to guide our model design. Nevertheless, the sign-symmetric structure of our model guarantees that not even side effects can cause such a bias. This underlines our proposal that the bias arises *outside* our model, namely in estimating the fixation distance which enters our model as the formal vergence parameters v_0. There are many potential cues (vertical disparity, accommodation, pictorial, etc.) for v_0, but none is generally reliable on its own. Thus, one expects that some stages outside our model take (weighted) averages over all cue signals *and a 'default' value*, such that when all cue signals drop out, one at least gets the correct percept at some sensible distance instead of at infinity or zero. Indeed, it is known (for review, Howard and Rogers 2002) that with many viewing-distance cues present, shape perception can be near-veridical (but not always: Todd and Norman 2003), whereas with minimal cues for viewing distance, a curvature bias appears that depends on distance, and that, in many subjects, goes though zero bias at distances of one to several meters. We conclude that this bias can reflect a sensible, robustness-motivated property of a neural stage which produces our formal v_0-parameter.

References

Anstis SM, Howard IP, Rogers B (1978) A Craik-Cornsweet illusion for visual depth. Vision Res 18:213–217

Aslin RN, Battaglia PW, Jacobs RA (2004) Depth-dependent contrast gain-control. Vision Res 44:685–693

Ben-Shahar O, Huggins PS, Izo T, Zucker SW (2003) Cortical connections and early visual function: intra- and inter-columnar processing. J Physiol 97:191–208

Berends EM, Liu B, Schor CM (2005) Stereo-slant adaptation is high-level, and does not involve disparity-coding. J Vision 5:71–80

van den Berg AV, van Ee R, Noest AJ (2005) Mixed visual reference frames: perceiving nonretino-centric visual quantities in a retino-centric frame. In: Rogowitz BE, Pappas TN, Daly SJ (eds) Human vision and electronic imaging X; Proceedings of SPIE vol 5666 pp 449–461

Bingham GP, Crowell JA, Todd JT (2004) Distortions of distance and shape are not produced by a single continuous transformation of reach space. Percept Psychophys 66:152–169

Burke WL (1985) Applied differential geometry. Cambridge University Press, Cambridge

Cumming BG (2002 An unexpected specialization for horizontal disparity in primate primary visual cortex. Nature 418:633–636

DeAngelis GC (2000) Seeing in three dimensions: the neurophysiology of stereopsis. Trends Cogn Sci 4:80–90

Domini F, Adams W, Banks MS (2001) 3D after-effects are due to shape and not disparity adaptation. Vision Res 41:2733–2739

Duke PA, Wilcox LM (2003) Adaptation to vertical disparity induced depth: implications for disparity processing. Vision Res 43:135–147

Erkelens CJ, Collewijn H (1985) Motion perception during dichoptic viewing of moving random-dot stereograms. Vision Res 25:583–588

Erkelens CJ, van Ee R (1998) A computational model of depth perception based on headcentric disparity. Vision Res 38:2999–3018

Gårding J, Porrill J, Mayhew JEW, Frisby JP (1995) Stereopsis, vertical disparity and relief transformations. Vision Res 35:703–722

Gillam B, Flagg T, Finlay D (1984) Evidence for disparity change as the primary stimulus for stereoscopic processing. Percept Psychophys 36:559–564

Glennerster A, McKee SP, Birch MD (2002) Evidence for surface-based processing of binocular disparity. Curr Biol 12:825–828

Gonzalez F, Perez R (1998) Modulation of cell responses to horizontal disparities by ocular vergence in the visual cortex of the awake macaca mulatta monkey. Neurosci Lett 245:101–104

Grossberg S, Swaminathan G (2004) A laminar cortical model for 3D perception of slanted and curved surfaces and of 2D images: development, attention, and bistability. Vision Res 44:1147–1187

von Helmholtz H (1867) Handbuch der physiologischen Optik. Vos, Hamburg

Hinkle DA, Connor CE (2002) Three-dimensional orientation tuning in macaque area V4. Nat Neurosci 5:665–670

Howard IP, Rogers BJ (2002) Seeing in depth, vol 2: Depth perception. Porteous, Toronto

Janssen P, Vogels R, Orban G (1999) Macaque inferior temporal neurons are selective for disparity-defined three-dimensional shapes. Proc Natl Acad Sci USA 96:8217–8222

Janssen P, Vogels R, Orban GA (2000) Three-dimensional shape coding in inferior temporal cortex. Neuron 27:385–397

Janssen P, Vogels R, Orban GA (2001) Macaque inferior temporal neurons are selective for three-dimensional boundaries and surfaces. J Neurosci 21:9419–9429

Johnston EB (1991) Systematic distortions of shape from stereopsis. Vision Res 31:1351–1360

Knapen T, van Ee R (2006) Slant perception, and its voluntary control, do not govern the slant aftereffect: multiple slant signals adapt independently. Vision Res, DOI:10.106/j.visres2006.03.027.

Koenderink JJ (1990) Solid shape. MIT Press, Cambridge

Koenderink JJ (1992) Fundamentals of bicentric perspective. Lecture notes in computer science, vol 653. Springer, Berlin Heidelberg New York, pp 233–251

Koenderink JJ (2003) Monocentric optical space. Lecture notes in computer science, vol 2756. Springer, Berlin Heidelberg New York, pp 689–696

Koenderink JJ, van Doorn AJ (1976) Geometry of binocular vision and a model for stereopsis. Biol Cybern 21:29–35

Koenderink JJ, van Doorn AJ (1987) Representation of local geometry in the visual system. Biol Cybern 35:367–375

Koenderink JJ, van Doorn AJ (1991) Affine structure from motion J Opt Soc Am A 8:377–385

Koenderink JJ, van Doorn AJ (2002) Image processing done right. Lecture notes in computer science, vol 2350. Springer, Berlin Heidelberg New York, pp 158–172

Koenderink JJ, Richards W (1988) Two-dimensional curvature operators. J Opt Soc Am A 5:1136–1141

Köhler W, Emery DA (1947) Figural after-effects in the third dimension of visual space. Am J Psychol 60:159–201

Lee B (1999) Aftereffects and the representation of stereoscopic surfaces. Perception 28:1155–1169

Lunn PD, Morgan MJ (1997) Discrimination of the spatial derivatives of horizontal binocular disparity. J Opt Soc Am A 14:360–371

Mayhew JEW (1982) The interpretation of stereo-disparity information: the computation of surface orientation and depth. Perception 11:387–403

Mayhew JEW, Longuet-Higgins HC (1982) A computational model of binocular depth perception. Nature 297:376–378

Nguyenkim JD, DeAngelis GC (2003) Disparity-based coding of three-dimensional surface orientation by macaque middle temporal neurons. J Neurosci 23:7117–7128

Nienborg H, Bridge H, Parker AJ, Cumming BG (2004) Receptive field size in V1 neurons limits acuity for perceiving disparity modulation. J Neurosci 24:2065–2076

Nishida S, Motoyoshi I, Andersen RA, Shimojo S (2003) Gaze modulation of visual aftereffects. Vision Res 43:639–649

Noest AJ (1994) Neural processing of overlapping shapes. In: Toet et al (eds) Shape in picture. NATO ASI series, vol 126, pp 383–392

Noest AJ, van Ee R, van den Berg, AV (2003) Disentangling retinal and head-centric disparity-coding involved in perception of metric depth from stereo. Perception 32(suppl):13–14

Ohzawa I (1998) Mechanisms of stereoscopic vision: the disparity energy model. Curr Opin Neurobiol 8:509–515

Petrov Y, Glennerster A (2004) The role of a local reference in stereoscopic detection of depth relief. Vision Res 44:367–376

Pianta MJ, Gillam BJ (2003) Monocular gap stereopsis: manipulation of the outer edge disparity and the shape of the gap. Vision Res 43:1937–1950

Pottmann H, Opitz K (1994) Curvature analysis and visualization for functions defined on Euclidean spaces or surfaces. Comput Aided Geom Des 11:655–674

Rogers BJ, Cagenello R (1989) Disparity curvature and the perception of three-dimensional surfaces. Nature 339:135–137

Rosenbluth D, Allman JM (2002) The effect of gaze angle and fixation distance on the responses of neurons in V1, V2, and V4. Neuron 33:143–149

Ryan C, Gillam B (1993) A proximity-contingent stereoscopic depth aftereffect: evidence for adaptation to disparity gradients. Perception 22:403–418

Sachs H (1987) Ebene isotrope Geometrie. Vieweg, Braunschweig

Steinman RM, Cushman WB, Martins AJ (1982) The precision of gaze. Hum Neurobiol 1:97–109

Strubecker K (1942) Differentialgeometrie des isotropen Raumes. III. Flächentheorie. Math Zeitschrift 48:369–427

Todd JT, Norman JF (2003) The visual perception of 3D shape from multiple cues: are observers capable of perceiving metric structure? Percept Psychophys 65:31–47

Trotter Y, Celebrini S, Stricanne B, Thorpe S, Imbert M (1992) Modulation of neural stereoscopic processing in primate area V1 by the viewing distance. Science 257:1279–1281

Trotter Y, Celebrini S, Stricanne B, Thorpe S, Imbert M (1996) Neural processing of stereopsis as a function of viewing distance in primate visual cortical area V1. J Neurophysiol 76:2872–2885

Trotter Y, Celebrini S (1999) Gaze direction controls response gain in primary visual-cortex neurons. Nature 398:239–242

van Ee R, Erkelens CJ (1996) Stability of binocular depth perception with moving head and eyes. Vision Res 36:3827–3842

van Ee R (2001) Perceptual learning without feedback and the stability of stereoscopic slant estimation. Perception 30:95–114

Wilson HR, Richards WA (1989) Mechanisms of contour curvature discrimination. J Opt Soc Am A 6:106–115