A new theory of structure-from-motion perception

Julian M. Fernandez
Institute for Sensory Research, Syracuse University, Syracuse, NY, USA

Bart Farell
Institute for Sensory Research, Syracuse University, Syracuse, NY, USA

Humans can recover 3-D structure from the projected 2D motion field of a rotating object, a phenomenon called structure from motion (SFM). Current models of SFM perception are limited to the case in which objects rotate about a frontoparallel axis. However, as our recent psychophysical studies showed, frontoparallel axes of rotation are not representative of the general case. Here we present the first model to address the problem of SFM perception for the general case of rotations around an arbitrary axis. The SFM computation is cast as a two-stage process. The first stage computes the structure perpendicular to the axis of rotation. The second stage corrects for the slant of the axis of rotation. For cylinders, the computed object shape is invariant with respect to the observer's viewpoint (that is, perceived shape doesn't change with a change in the direction of the axis of rotation). The model uses template matching to estimate global parameters such as the angular speed of rotation, which are then used to compute the local depth structure. The model provides quantitative predictions that agree well with current psychophysical data for both frontoparallel and non-frontoparallel rotations.

Keywords: structure from motion, kinetic depth effect, affine structure, shape perception

Citation: Fernandez, J. M., & Farell, B. (2009). A new theory of structure-from-motion perception. *Journal of Vision, 9*(11):23, 1–20, http://journalofvision.org/9/11/23/, doi:10.1167/9.11.23.

Introduction

Our current understanding of how motion triggers the perception of three-dimensional structure comes from the data of a special case: Structure from motion (SFM) has been studied almost exclusively from rotations about an axis in the frontoparallel plane. Perceived shape in this condition is usually non-veridical. This has given rise to two main theoretical approaches:

1. SFM from 1st-order velocities: One theory is that humans recover SFM from the 1st-order velocity field. This implies that shape is recoverable only up to a scaling factor in depth (Norman & Todd, 1993; Todd, 1998; Todd & Norman, 1991; Werkhoven & van Veen, 1995). Hence, accuracy is low for judgments requiring veridical perception of Euclidean metric structure, such as judgments of lengths or angles (Braunstein, Liter, & Tittle, 1993; Cornilleau-Pérès & Droulez, 1989; Eagle & Blake, 1995; Hogervorst, Kappers, & Koenderink, 1993; Liter, Braunstein, & Hoffman, 1993; Norman & Lappin, 1992; Todd & Bressan, 1990). Conversely, accuracy is high for judgments of an object's "affine"[1] structure—the structure up to a scaling factor in depth—such as depth order between pair of points, parallelism between lines defined by pairs of points on a planar surface, and coplanarity among points (Braunstein et al., 1993; Eagle & Blake, 1995; Hogervorst et al., 1993; Liter et al., 1993; Tittle et al., 1995; Todd & Bressan, 1990). Koenderink and van Doorn (1991) developed one of the best-known theoretical approaches to computing affine structure from the first-order optic flow.

2. SFM from 2nd-order optic flow and perspective effects: There is evidence that humans can use second-order optic flow information (that is, accelerations) and perspective information in computing the SFM. Regarding perspective effects, Eagle and Hogervorst (1999) found that shape discrimination thresholds decreased with stimulus size under perspective projection, but under orthographic projection, thresholds increased. Regarding the effects of accelerations, Hogervorst and Eagle (1998, 2000) showed that errors in the estimate of speeds and accelerations could explain some pattern of misperceptions in SFM.

3. Surface slant from def: A more radical view is that of Domini et al. (Domini & Braunstein, 1998; Domini & Caudek, 2003; Domini, Caudek, & Richman, 1998). They propose that the perceived slant of a surface is a function of the 1st-order optic-flow property *def* and predict that the structure of a shape recovered from SFM will be internally inconsistent. The authors suggest that this prediction is in agreement with numerous experimental results (Domini & Braunstein, 1998; Domini et al., 1998).

There is an alternative that is also in agreement with experimental results and that results in internally consistent shapes. One of the aims of this article is to develop this alternative theory. The second aim is to make the theory general and independent of the axis of rotation.

doi: 10.1167/9.11.23 Received December 10, 2008; published October 26, 2009 ISSN 1534-7362

Previous models

Many models for computing SFM have been proposed. Some of them attempt to describe how humans perceive SFM; others show how SFM can be computed from the available information. Among the latter, the best known are the class of models based on Ullman's theorem (and its variants) that under orthographic projection, three views of four non-coplanar points from a rigid object are sufficient for uniquely determining the 3D structure of the points (up to a mirror reflection) (Ullman, 1979). Ullman later proposed an "incremental rigidity algorithm" for recovering structure from input data over time (Ullman, 1984). While these studies are extremely interesting in their own right, they do not tell us how neurons in the brain could solve the problem. In fact, there is no evidence that visual neurons could explicitly track positions of fine image features over time, as implied by the input representation of these models.

As we previously mentioned, more recent psychophysical data suggest that humans use velocity information instead of positional locations of image features for surface interpolation in SFM tasks (Treue et al., 1995). This result led to a major modification of the incremental rigidity algorithm (Hildreth, Ando, Andersen, & Treue, 1995). One problem with most of the earlier models, however, is that they compute the veridical Euclidean metric structure of the object as the output (Hildreth et al., 1995; Ullman, 1979, 1984), which contradicts the psychophysical data previously mentioned. An exception is the model previously mentioned of Koenderink and van Doorn (1991), which computes affine structure from the first-order optic flow.

Non-frontoparallel rotations

Most of the theoretical approaches mentioned so far are based on data from frontoparallel rotations, and their scope is limited to that particular condition [the exception is the theory of Koenderink and van Doorn (1991), which applies not only to non-frontoparallel rotations, but also to semi-parallel projections and movement in depth, two topics not covered by our theory]. Until recently, few empirical studies looked at non-frontoparallel axes of rotation. A first study, by Loomis and Eby (1988), found that the perceived depth of elongated ellipsoids becomes a smaller fraction of the simulated depth the more the angle of rotation deviates from frontoparallel. A second study, by Liter et al. (1993), reached the same conclusion using 3D "objects" made with five randomly positioned dots.

There are other studies involving rotations about slanted axes, but these did not examine the perception of shape from motion per se, but instead looked at other issues, such as the perceived inclination of the axis of rotation (Domini & Caudek, 1999; Pollick, Nishida, Koike, & Kawato, 1994), or the perception of rigidity (Todd & Bressan, 1990).

More recently, we investigated in greater detail the ability of humans to recover shape from rotations about a slanted axis (Fernandez & Farell, 2007). We found that in this condition the perceived structure is *not affine*. Moreover, violations of affinity can result in perceived *depth-order violations* along the line of sight. We found, for cylinders rotating about its longitudinal axis, that as the axis of rotation changes its inclination, shape constancy is maintained through a trade-off: Humans perceive a constant object shape relative to a changing axis of rotation and they do this by introducing an inconsistency between the perceived speed of rotation and the 1st-order optic flow. Shape constancy demands this inconsistency because observers do not perceive the inclination of the axis of rotation veridically. The observed depth-order violations are the cost of the trade-off. Note that objects other than cylinders do not show this kind of shape constancy. As mentioned above, Loomis and Eby (1988) found that the perceived shape of rotating ellipsoids changes with changes in the slant of the axis of rotation.

Thus, structure recovered from rotations around an axis in the frontoparallel plane preserves affine structure. The more general case of structure recovered from rotations around an arbitrary axis does not preserve affinity. But, for cylinders rotating about its longitudinal axis, it does preserve perceived shape relative to the perceived axis of rotation.

Here we show that recovering the depth structure in the general case can be decomposed into two stages. In the first stage the depth structure relative to the axis of rotation is computed: the information about this depth structure is available from the optic flow field in the projected component of the speeds perpendicular to the axis of rotation. This computational problem is mathematically similar to computing the depth structure for frontoparallel rotations. The second stage computes a correction that takes into account the slant of the axis of rotation. This gives the depth structure for an arbitrary axis of rotation slanted with respect to the frontoparallel plane.

Thus, there are currently two major issues that need to be understood and quantified in order to develop a general model of the SFM computation. The first is how to characterize the algorithm for recovering structure from frontoparallel rotations (current theories lack an algorithm that allows one to predict the recovered depth structure even in the special case of a frontoparallel axis of rotation). The approach usually taken previously has been to describe the computation behind the *depth-scaling* parameter, that is, the ratio between perceived and simulated depth. This approach assumes that the visual system computes affine structure first. Then, in a second step, a scaling factor is "assigned." This transforms the initially computed affine structure into a metric object. But there are other ways to think about this problem of computing SFM, and in this article we used a different approach.

The second requirement is to characterize the SFM computation for the general case of non-frontoparallel rotations.

Our purpose here is to develop a theory that answers these two basic issues.

The model

The model's structure is conceptually simple, and consists of two parts. One part is a template-matching mechanism that uses the optic-flow field as the input to compute two global parameters: the perceived speed of rotation and the perceived gradient of the speeds perpendicular to the frontoparallel projection of the axis of rotation. The second part consists of an operation that generalizes the structure-from-motion computation from the case of a frontoparallel axis of rotation to the general case of arbitrary non-frontoparallel axes.

Recovering depth structure

Frontoparallel rotations

The depth structure for frontoparallel rotations is linearly related to the velocity field (see, for instance, Fernandez, Watson, & Qian, 2002):

$$\frac{\Delta Z}{Z_0} = -\frac{\Delta v_x}{\Omega}, \tag{1}$$

where Z_0 is the distance to the object, and Ω is the angular speed of rotation. Equation 1, which implicitly assumes rigidity, gives the difference in depth ΔZ (positive towards the observer) between any two points on the object from their differences in horizontal retinal velocity Δv_x. Equation 1 assumes, without loss of generality, a vertical axis of rotation.

If SFM is recovered only from the first-order optic flow, as psychophysical data seems to show (Norman & Todd, 1993; Todd, 1998; Todd & Norman, 1991; Werkhoven & van Veen, 1995), then Ω cannot be recovered veridically.[2] As a result, the depth structure can be recovered only up to a scaling factor in depth. Current theories lack an algorithm for predicting this scaling factor, which is needed to predict the recovered depth structure in the special case of a frontoparallel axis of rotation. Computing the scaling factor is mathematically equivalent to computing the perceived angular speed of rotation, Ω_{obs}, which can be used in Equation 1 instead of Ω, the actual speed of rotation. Thus, using Ω in Equation 1 allows recovery of the veridical shape of the object, whereas using Ω_{obs} allows recovery of the perceived shape.

Note that if Ω_{obs} is non-veridical, in principle the object should end up looking either non-rigid or as a texture flowing over the surface of an otherwise rigid object. Perceptually, some cases (a minority) give the appearance of a texture flowing over the surface of an otherwise rigid object. But in all cases the objects appear rigid. One reason could be that the object's non-rigidity is below threshold. A big factor in discriminating rigidity and non-rigidity depends on temporal factors—memory—rather than just the magnitude of shape change. Large changes in the object can remain undetected because we do not keep a faithful 3D representation of the object in memory across a sufficiently long time, but only what amounts to an instantaneous one. In a task in which a 3D representation of the object had to be kept in memory, then the non-rigidity would perhaps become "visible." Data on rigidity-versus-non-rigidity discriminations show that non-rigid objects can be perceived as rigid and vice versa, depending on whether specific optic-flow properties change, or do not change, over time.[3] As an example, Hogervorst, Kappers, and Koenderink (1997) found that non-rigid displays were not discriminable from rigid ones when every set of three consecutive views in the sequence had a rigid solution. They concluded that observers combined only a few views at a time. That is, they used a temporal local description of the flow. Only very large changes in structure were detected over large changes in rotation.

Non-frontoparallel rotations

We recently showed (Fernandez & Farell, 2007) that when the axis of rotation is not in the frontoparallel plane, the 3D structure of the object under orthographic projection must satisfy the following two relations:

$$\frac{\Delta Z}{Z_0} = -\frac{\Delta v_x}{\Omega \cos\theta} + \Delta y \tan\theta, \tag{2}$$

and

$$\Omega = -\frac{\partial_x v_y}{\sin\theta}, \tag{3}$$

so that

$$\frac{\Delta Z}{Z_0} = \left(\frac{\Delta v_x}{\partial_x v_y} + \Delta y\right)\tan\theta, \tag{4}$$

where θ is the actual of inclination of the axis of rotation ($0° < \theta < 90°$) from the frontoparallel plane, and $\partial_x v_y$ is the derivative with respect to x (horizontal direction) of the vertical component of the retinal velocity, v_y. Note that $\partial_x v_y$ is constant across the image and thus independent of depth (see Fernandez & Farell, 2007, for a demonstration). Equation 4, which implicitly assumes rigidity, gives the

difference in depth ΔZ (positive towards the observer) between any two points on the object from their differences in horizontal retinal velocity Δv_x and angular vertical position Δy. It is assumed here that the axis of rotation lies within the sagittal plane passing through the eye. This choice does not represent a loss of generality, but is a direct consequence of assuming that the axis of rotation projects into a vertical line in the frontoparallel plane (see Equations 1 and 4). If the projection of the axis of rotation onto the frontoparallel plane were not vertical, Equations 1 and 4 would still be similar, but v_x would be replaced by the speeds perpendicular to the frontoparallel projection of the axis of rotation, and v_y by the speeds parallel to this projection (also, Δy would be measured in the direction parallel to the projection of the axis).

If SFM for non-frontoparallel rotations were recovered only from the first-order optic flow, then θ could not be recovered veridically (this is in agreement with psychophysical data, see for instance Fernandez & Farell, 2007). Then depth structure could be recovered only up to a scaling factor in depth by using the perceived angle of inclination, θ_{obs}, in Equation 4.

However, we showed (Fernandez & Farell, 2007) that the perceived depth structure does not follow Equation 4, but rather satisfies the following relation:

$$\frac{\Delta Z}{Z_0} = \left(\lambda \frac{\Delta v_x}{\partial_x v_y} + \Delta y\right) \tan\theta_{\text{obs}}, \tag{5}$$

where the factor λ accounts for the departure of the perceived angular speed of rotation, Ω_{obs}, from optic-flow consistency.[4]

To retrace the picture we've sketched so far, there is, first of all, the speed of rotation in the stimulus, Ω, that satisfies Equation 3. If the axis slant, θ, is perceived non-veridically as θ_{obs}, it is still possible to use Equation 3 by substituting θ_{obs} for θ. The result is a non-veridical perceived speed of rotation that is consistent with the first-order optic flow. This is equivalent to recovering the shape from Equation 5 using $\lambda = 1$. Having $\lambda \neq 1$, in agreement with psychophysical data, implies that Ω_{obs}, the perceived speed of rotation, is not consistent with the first-order optic flow.

As shown in Figure 1, the geometrical effect of λ is to multiply all distances relative to the axis of rotation rather than to the frontoparallel plane. The direction of this scaling is perpendicular to the frontoparallel plane, so that the retinal image will remain unchanged. In this sense, the recovered structure of the object is not affine. Let us characterize this recovered structure as λ-affine.

λ can be viewed as a correction factor whose purpose is to allow the observer to perceive the stimulus shape as constant as the axis of rotation changes. In such a case λ could in principle be computed from the optic flow (in Fernandez & Farell, 2007, we showed that this is possible). However, this way of thinking about λ has the

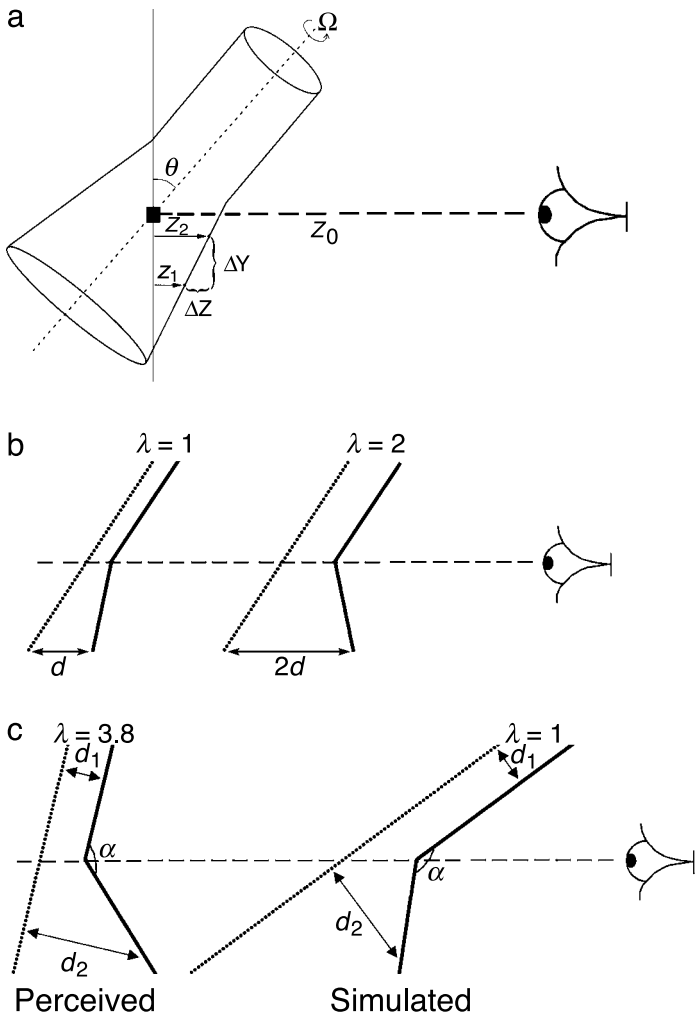

Figure 1. Non-affine structure. a) Schema showing experimental stimulus. b) Effect of changing λ for a constant inclination of the axis of rotation. Side view (sagittal cut) from the object shown in (a). Changing λ result in a depth order reversal of the bottom half, as all the distances to the axis of rotation (dotted line) double. c) Same as (b), but this time showing the difference between the perceived and simulated object. The shape of the perceived object can be made independent of the inclination of the axis of rotation by using an adequate non-unity value of λ, leading also to depth-order violations, as seen here. By contrast, a value of $\lambda = 1$ gives an affine structure, preserving depth order but not the values of d_1, d_2 and α.

disadvantage of imposing an additional computational burden on the visual system. Alternatively, λ can be viewed as a bias intrinsic to the observer rather than being computable from the input. We will follow this latter interpretation here. In addition, we propose that λ is a constant.

A constant value for λ results in additional constraints. For instance, for cylinders, we shown in Appendix A that λ can be a constant only if $\sin\theta_{\text{obs}} \propto \sin\theta_{\text{sim}}$ (the constraint comes from the psychophysical fact that, for cylinders rotating about its longitudinal axis, the perceived

stimulus shape does not changes with changes in the axis of rotation). Now, by definition, we have (see Appendix A):

$$\sin\theta_{\text{obs}} = \frac{-\partial_x v_y}{\lambda\ \Omega_{\text{obs}}}. \quad (6)$$

The analogy between Equations 3 and 6 suggests a more precise interpretation of $1/\lambda$ as the proportionality factor between perceived and simulated gradient. We will call the perceived gradient ∇_{obs} ($=\partial_x v_y/\lambda$). We can rewrite Equation 5 in a format that explicitly reduces to Equation 1 (i.e., the frontoparallel case) when $\theta = 0$ (and thus, $\nabla_{\text{obs}} = 0$). Using Equations 5 and 6 we get, after some work:

$$\frac{\Delta Z}{Z_0} = -\frac{\Delta v_x}{\sqrt{\Omega_{\text{obs}}^2 - \nabla_{\text{obs}}^2}} + \frac{\Delta y}{\sqrt{(\Omega_{\text{obs}}/\nabla_{\text{obs}})^2 - 1}}. \quad (7)$$

Equation 7 shows the two stages mentioned in Introduction. The first term in the right side of Equation 7 represents the frontoparallel structure (up to a scaling factor in depth), or equivalently, the structure relative to the axis of rotation. The second term is a correction that takes into account the slant of the axis of rotation, becoming zero for frontoparallel rotations.

Equation 7 allows us to compute the relative depth structure of an arbitrary rotating object from the two parameters Ω_{obs} and ∇_{obs}. These parameters can be obtained from template matching, as described below in The template-matching model.

The template-matching model

Template matching plays a fundamental role in our SFM model. Template matching implements what we hypothesize is the contribution of area MST to the SFM computation (a contribution that is made by only a subset of MST cells). Many MST cells are tuned to optic-flow patterns that are useful for heading perception, and successful template-matching models of cells with these tuning properties have been developed (Perrone & Stone, 1998). Alternative models of MST use subsets of neurons whose properties are better for modeling the perception of biological motion (Giese & Poggio, 2003). Overall, the evidence suggests that MST contains subsets of cells tuned to different global patterns of motion, each of which is best suited to perform one of various classification or recognition tasks, e.g., determining heading direction, recognizing biological motion, classifying shape from motion. Our model MST cells have tuning properties that include these cases.

For illustrative purposes, let us first consider the case where all velocities are parallel; this happens, for instance, for frontoparallel rotations. To define the tuning properties of our MST model cells, we define for each cell the matching variable I:

$$I = \iint [\mathrm{f}(x,y) - \mathrm{g}(x,y)]^2 \mathrm{d}x\mathrm{d}y. \quad (8)$$

I is a measure of the difference between the input velocity pattern, f(x, y), and the template velocity pattern, g(x, y), where (x, y) gives the spatial retinal coordinates in the image plane. I takes its minimum value when both template and input patterns are the same. The response, r, of the model MST cell is then defined as

$$r = e^{-\beta I}, \quad (9)$$

where β is a constant.

Thus, r is a maximal response when the template and input velocity patterns are the same. Equation 8 is valid when the functions f and g are scalar quantities (which is the case for frontoparallel rotations, where all velocities are parallel). In the more general case of non-frontoparallel rotations, f and g are vectors representing the velocity at any point in the retinal image. These velocities could have any direction. To generalize Equation 8 to vector variables we modify it in the following way.

$$I = \iint \left[\vec{\mathrm{f}}(x,y) \cdot \frac{\vec{\mathrm{g}}(x,y)}{|\vec{\mathrm{g}}(x,y)|} - |\vec{\mathrm{g}}(x,y)|\right]^2 \mathrm{d}x\mathrm{d}y$$
$$= \iint \left[\mathrm{f}(x,y) - \mathrm{g}(x,y)\right]^2 \mathrm{d}x\mathrm{d}y, \quad (10)$$

where the dot represents the scalar product, and the arrows indicate that the variables are vectors. The second equality represents a simplified notation of the first, where:

$$\mathrm{f}(x,y) = \vec{\mathrm{f}}(x,y) \cdot \frac{\vec{\mathrm{g}}(x,y)}{|\vec{\mathrm{g}}(x,y)|} \text{ and } \mathrm{g}(x,y) = |\vec{\mathrm{g}}(x,y)|, \quad (11)$$

thus f and g are scalars, with f being the component of the input speed in the direction of the speed in the template, and g being the speed (modulus) in the template, regardless of direction. The modification has the following effect: rather than comparing the whole 2D speed input pattern to the template, it only compares, at each retinal position, the component of the speed in the input which is parallel to that in the template.

To make the problem easier to analyze and keep the number of templates low, we will only use templates that

have speeds in the horizontal direction (this, of course, does not imply a loss of generality; the treatment that follow could also include templates representing slanted objects). This restriction will ensure that only the horizontal components of the retinal speeds are matched. For cylinders only, this implies that the same template will match a given rotating cylinder regardless of the slant of the axis of rotation. This works because when the axis of rotation changes its slant, only the vertical components of the retinal speed change. For objects other than cylinders,[5] the velocity field is not independent of the axis slant.[6] As a consequence of this, perceived shape will change if the axis slant changes. This is consistent with the psychophysical data of Loomis and Eby (1988) on rotating ellipsoids.

Are such templates physiologically plausible? Most models of MST tuning use a tiled input from MT-like neurons, each tuned to a particular speed and direction. The templates built in this way will care about the local input velocity, not just the component of the speed in one direction. To build our MST template cell, we need to make use of complex V1-like cells. Because they suffer from the aperture problem, these cells are not selective for stimulus velocity; rather, they are selective only for the component of velocity orthogonal to their preferred spatial orientation (Simoncelli & Heeger, 1998). Physiologically, it is plausible that some MST cells use complex-cell V1 input, either received directly or inherited via area MT. Though this remains speculative at this point, it delivers a considerable computational advantage for problems like SFM.

Template matching for the angular speed of rotation

Let us first analyze the simpler case of frontoparallel rotations. It is not difficult to show that at each location (x, y) the optic-flow pattern from an object rotating about a frontoparallel axis can be written as the product of a shape factor and the angular speed of rotation, Ω (see Appendix A). Thus $f(x, y) = \Omega F(x, y)$, where F is a function that depends only on the object's shape (as expressed either by the stimulus or the template). This is a function not only of position (x, y), but also of parameters that define the shape (we will keep the treatment general at this point, leaving for later the specification of the template shapes used in our simulations). For instance, a rotating elliptical cylinder will have three parameters, S, C and φ, where S represents the size, C represents the ellipticity, and φ the angle between the major axis and the line of sight. The specific set of functions used in making the templates is described later in the subsection A minimalist template set.

If the axis of rotation is slanted, the previous paragraph still holds true for cylinders, but it is no longer valid for other shapes. In the general case, v_x is still proportional to Ω, but instead of a shape factor we have a more complex expression that is also a function of the axis slant (see Appendix A).

Let us represent by α and γ_0 the sets of parameters representing the shapes of the stimulus and the template, respectively, and Ω and Ω_0 the angular speeds of stimulus and template, respectively. It is easy to show (replacing $f(x, y) = \Omega F(x, y, \alpha)$ and $g(x, y) = \Omega_0 G(x, y, \gamma_0)$ in Equation 8)[7] that the matching variable I, defined in Equation 8, becomes (remember that this is only for frontoparallel rotations):

$$I = A_G(\gamma_0)\left[\Omega_0 - \Omega\frac{B(\alpha,\gamma_0)}{A_G(\gamma_0)}\right]^2 + \Omega^2\left[A_F(\alpha) - \frac{B^2(\alpha,\gamma_0)}{A_G(\gamma_0)}\right], \quad (12)$$

where

$$\begin{aligned} A_F(\alpha) &= \iint F^2(x,y,\alpha)\mathrm{d}x\mathrm{d}y; \\ A_G(\gamma_0) &= \iint G^2(x,y,\gamma_0)\mathrm{d}x\mathrm{d}y; \\ B(\alpha,\gamma_0) &= \iint F(x,y,\alpha)G(x,y,\gamma_0)\mathrm{d}x\mathrm{d}y. \end{aligned} \quad (13)$$

If we have a population of neurons representing different flow templates (that is, having different values of Ω_0 and γ_0), we can see from Equations 9 and 12 that, for a fixed γ_0, the population response will have a bell-shaped curve. The peak of this curve (which is always a maximum because A_G is always positive) is a function of the stimulus angular speed (Ω) and shape (represented by the set of parameters α). Because the population of cells contains cells tuned to various values of γ_0, we will have many cells that respond maximally. The tuning parameter Ω_0 of the maximally responding cell (for a given γ_0) will be:

$$\Omega_0 = \Omega\frac{B(\alpha,\gamma_0)}{A_G(\gamma_0)}. \quad (14)$$

Thus, cells will respond maximally along a "curve"[8] in parameter space (Ω_0, γ_0). The average value of Ω_0 obtained from the population response will then be:

$$\langle\Omega_0\rangle = \Omega\left\langle\frac{B(\alpha,\gamma_0)}{A_G(\gamma_0)}\right\rangle. \quad (15)$$

We hypothesize that the relationship between Ω_0 and the perceived angular speed of rotation, Ω_{obs}, is simple and direct. For instance, if Ω_{obs} were proportional to $\langle\Omega_0\rangle$, then Equation 15 tells us that the perceived rotational speed will be proportional to the simulated rotational speed.[9] The constant of proportionality will be different

for different shapes and will depend on the particular family of templates used. This is a highly desirably property that is very difficult, if not impossible, to attain with simpler models that treat Ω_{obs} as some function of *def* or other optic-flow properties that are not shape invariant.

Generalization to slanted axes of rotation. Having started with the more intuitively accessible case of a frontoparallel axis of rotation, we are now in a position to generalize to the case where the axis of rotation is slanted. Equations 12 and 15 can be generalized to slanted axes, because the optic flow can be decomposed into a product of Ω and a "shape" factor. For slanted axes, the "shape" factor is a function of the shape and the slant of the axis of rotation (see Appendix A.2.2). That means that the previous derivations are not restricted to frontoparallel rotations but they are also valid for slanted axes of rotation. Note also that for non-frontoparallel rotations, for shapes other than cylinders, the constant of proportionality in Equation 15 will be a function of the axis slant—because $B(\alpha, \gamma_0)$ is a function of the axis slant—but at any slant Ω_{obs} will still be proportional to Ω.

The perceived scaling factor in depth is just the ratio between the simulated and perceived Ω s. Equation 15 shows that, if Ω_{obs} is proportional to $\langle\Omega_0\rangle$, this ratio will be constant for a given shape and axis slant, and, in general, different for different shapes and different axis slants (although remember that for cylinders it will not change with the axis slant). This is in agreement with psychophysical data, as will be shown in Results.

Template matching for gradients

As Equation 7 shows, to compute the relative depth structure of an arbitrary rotating object we need two parameters, Ω_{obs} (the perceived speed of rotation) and ∇_{obs} (the perceived gradient). We showed in the previous section how Ω_{obs} could be obtained by template matching. Now we will show how to obtain ∇_{obs} from template matching.

Our hypothesis is that the gradient $\partial_x v_y$ is computed by template matching. Computing a gradient for rotations using templates is actually much simpler than computing Ω_{obs}. This is because for rotations $\partial_x v_y$ is constant across the image, so for any rotating object we have $v_y = x\partial_x v_y$. We will abbreviate $\partial_x v_y$ as ∇, so $v_y = x\nabla$.

As before, we assume that our templates will only match the component input velocity that is parallel to that in the template. This direction will be the same across the entire template. More explicitly, we assume that the templates are described by $g(x, y) = \nabla_0 G(x)$ where ∇_0 is a constant and $G(x)$ is an arbitrary function of x. We assume also that the family of templates includes all motion directions, which are obtained by rotating the template just defined.

Because the templates only care about the speeds in their preferred directions, we can use as the input only the speed component of the input parallel to that in the template. For an object rotating about an axis whose projection onto the frontoparallel plane is vertical, we can write $f(x, y) = x\nabla$. Note again that the gradient ∇, and thus the input, is independent of the object's shape (see Equation 3).

By analogy to Equation 14, the tuning parameter ∇_0 of the maximally responding cell will be:

$$\nabla_0 = \nabla \frac{B}{A_G}, \tag{16}$$

where (see Equation 13):

$$\begin{aligned} A_G &= \iint G(x)^2 \, dx dy; \text{ and} \\ B &= \iint x G(x) dx dy. \end{aligned} \tag{17}$$

It is clear that the gradient "perceived" by the templates (i.e., ∇_0) will be proportional to the simulated gradient, regardless of the stimulus shape and regardless of the particular family of templates, i.e., the specific form of $G(x)$. As an example, let us assume $G(x) = \eta x$, where η is a constant. Then, it follows that $\nabla_0 = \nabla/\eta$. This implies, by the definition of λ, that $\eta = \lambda$.

A minimalist template set

It is important to realize that template matching can be used to model psychophysical data to any degree of accuracy. This can be done just by using as many templates as stimuli one wants to model. For each stimulus $F(x, y, \alpha)$ one could use Equation 15 to obtain $B(\alpha, \gamma_0)$ and then Equation 13 to find the corresponding template $G(x, y, \gamma_0)$. It is more interesting and productive to take the opposite approach and ask: How well can we explain the available psychophysical data by keeping the number of templates and their complexity to a minimum?

In what follows, we assume that there are only two families of templates: the first one for detecting velocity fields in which curvature is absent (e.g., planar surfaces), and the second one for detecting objects in which curvature is present (e.g., cylinders). These can be considered as a set of first- and second-order filters. We assume that a stimulus generates a population response in which each filter contributes a value for its preferred parameter (either Ω_0 or ∇_0) to the computation of the output parameter. The computation is a weighted average in which each filter parameter has a weight proportional to the intensity of its filter's response.

Two sets of templates, one for planar surfaces and one for cylinders, were used to obtain the data shown in

Results. Remember that a template is the product of a shape factor (or function) and an angular speed of rotation, Ω_0.

Planar Surfaces: The first set of templates consists of the product of one specific planar surface (that is, one having a fixed slant and tilt) with various angular speeds of rotation. Thus, a specific template within this family is identified by just one parameter, the angular speed of rotation (see Methods).

Cylinders: The second family of templates consists of the product of an elliptical cylinder rotating about its longitudinal axis, with various angular speeds. The elliptical cylinders could have different ellipticities, but are all similarly oriented, that is, the major axis of each ellipse makes the same fixed angle γ_0 with the line of sight. A specific template within this second family is identified by two parameters: the ellipticity and the angular speed of rotation (see Methods). In addition, we scaled the stimuli to a standard size that matches that of the templates. An alternative but equivalent option would have been to assume that templates come in families with size as an additional parameter, such that the cylinder templates and the stimuli match in size.

This limited set of templates is, of course, an over-simplification. If the visual system uses templates, it would likely possess a far larger variety that our two sets—for instance, planar templates having various tilts and slants rather than a single tilt and a single slant; and cylinders with the major axis of the ellipse forming various angles to the line of sight. However, our goal here is to show how well our minimalist set of templates reproduces psychophysical data in spite of its simplicity. This will demonstrate the explanatory power of the template matching approach.[10]

Templates and their relationship to perceived biases

An incomplete set of templates is one source of the biases predicted by the model. We believe it is unrealistic to assume that the visual system possesses a complete set, a set than spanned all object forms, orientations, sizes, rotational speeds, tilts and slant of the axis of rotation, and so forth. The number of template cells would be problematic. Moreover, this number would assure that observers would not have previously experienced many of the corresponding stimuli, raising the question of the origins of templates representing them. In any case, our intention is not to equate the templates we employ in the model with those that plausibly exist in the human brain. Instead, we want to show that our model's approach and architecture can work, and can work without cherry-picking the templates, or assuming that all possible templates are available.

There are other ways aside from an incomplete set of templates for perceptual biases to arise. For instance, even if all possible templates were available, a bias would result from an imperfect mapping between the template's tuning parameter and the circuitry used to compute the depth structure.

Results

The model makes a number of predictions about the perception of rotating objects. We evaluate these predictions below under four headings: (1) perceived speed of rotation, (2) perceived slant of planar surfaces, (3) perceived curvature of elliptical cylinders and (4) perceived slant of the axis of rotation. The selected set of psychophysical data that we choose to model constitutes, to the best of our understanding, a representative sample of the available quantitative psychophysical data on rotating objects.

Perceived speed of rotation

We assume that the set of templates for planar surfaces specifies a planar surface with fixed slant, σ_0, and tilt, τ_0, and various rotational speeds. Then Equation 14, which gives the speed tuning parameter Ω_0 of the maximally responding cell[11] becomes (see Appendix A):

$$\Omega_0 = \Omega \frac{\sigma}{\sigma_0} h(\tau, \tau_0) \propto def, \tag{18}$$

where Ω, σ and τ are the speed of rotation, slant and tilt of the stimulus, respectively; and h is a function of τ and τ_0. The rightmost part of Equation 18 is a consequence of the fact that $def = \Omega\sigma$ for planar surfaces. If the perceived speed of rotation, Ω_{obs}, is proportional to Ω_0 then, by Equation 18, Ω_{obs} is proportional to def. This agrees with the psychophysical data of Domini, Caudek, and Proffitt (1997) (see their Figure 7).

However, in subsequent experiments Domini and Caudek (1999) found that Ω_{obs} is proportional to $def^{1/2}$ (see their Figure 15, reproduced here in Figure 2b). It is unclear why the result differed in the two cases, although, due to error uncertainties, either set of data could probably be fitted by a linear function, a square root function, or any function in between.

Let us now assume that the latter experiments where Ω_{obs} is proportional to $def^{1/2}$ are correct. If our model is going to predict this, then we need to assume that the relationship between Ω_{obs} and Ω_0 is not linear; rather, we will assume that Ω_{obs} is proportional to $\Omega_0{}^{1/2}$. One reason why the visual system might prefer a non-linear relationship between the cell's tuning parameter Ω_0 and Ω_{obs} is a trade-off: for a linear relationship the perceived slant, σ_{obs}, would be independent of the surface slant, σ (see next section). A non-linearity overcomes this problem.

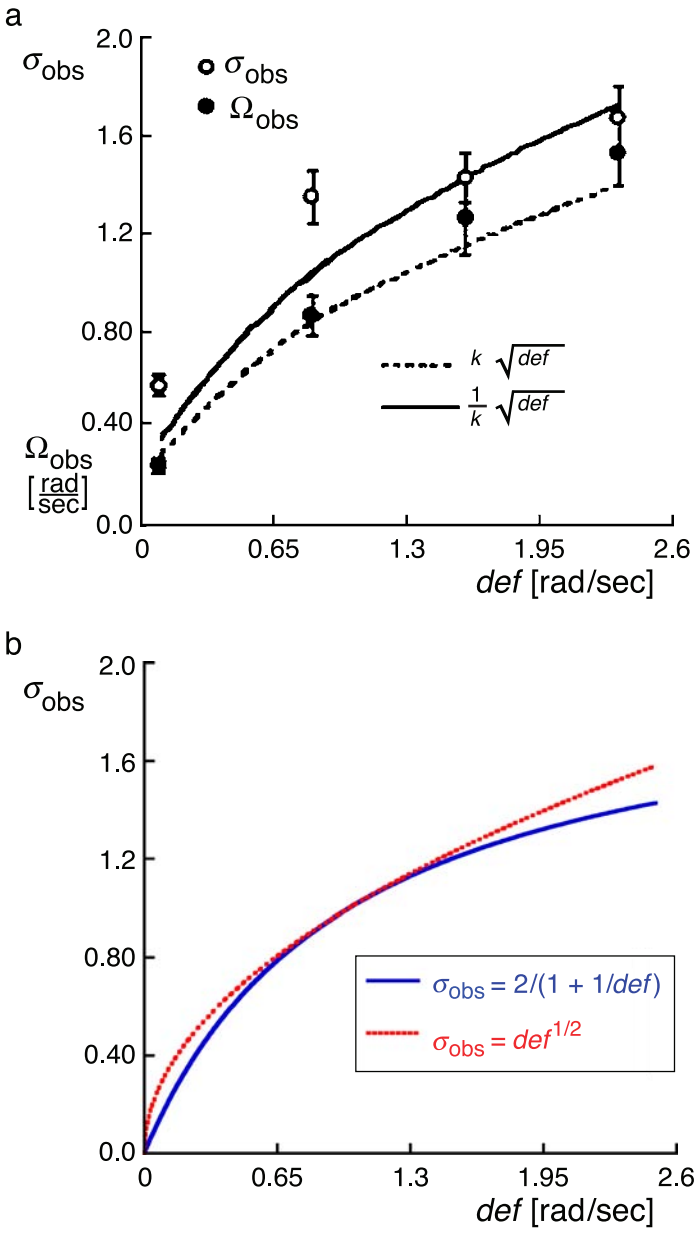

Figure 2. a) The perceived slant of a planar surface, σ_{obs}, as a function of *def,* from Figure 15 of Domini and Caudek (1999). (Reproduced with permission). Notice that the vertical axis has two labels. b) Model predictions for the conditions used in Domini and Caudek (1999). For the range of values of *def* tested, both Equations 19 and 20 provide a reasonable fit to the data. Same results (not shown) are obtained for perceived angular speed of rotation.

The non-linearity $\Omega_{obs} \propto \Omega_0{}^{1/2}$ results in $\Omega_{obs} \propto def^{1/2}$, which is consistent with the results of Domini and Caudek (1999), as shown in Figure 2a. Note that assuming the above nonlinearity results in specific model predictions for the behavior of σ_{obs}. As shown in the next section, σ_{obs} is now predicted to behave as $\sigma_{obs} \propto \sigma^{1/2}$, which is also consistent with psychophysical data (Domini & Caudek, 1999).

Perceived slant of planar surfaces

The perceived slant of a planar surface, σ_{obs}, is a function of its perceived rotational speed, Ω_{obs}. They are related, by definition, through the relationship: $def = \Omega\sigma = \Omega_{obs}\sigma_{obs}$ (see Appendix A). This expression is valid for frontoparallel rotations only (as far as we know). Thus, the model predictions that follow in this section are restricted to frontoparallel rotations. This is enough for our purposes because the data from Domini and Caudek (1999) that we are going to model was also restricted in the same way.

It is not difficult to show using $\sigma_{obs} = def/\Omega_{obs}$ and Equation 18 that, if Ω_{obs} is proportional to Ω_0 (i.e., $\Omega_{obs} = k\Omega_0$), then $\sigma_{obs} = \sigma_0/[k\ h(\tau, \tau_0)]$. Thus, σ_{obs} is independent of both the simulated slant, σ, and *def*. This disagrees with psychophysical data (see Figure 15 of Domini & Caudek, 1999, reproduced here in Figure 2a). There are two ways to modify the model to bring it in line with the psychophysical data. One way is to assume that our choice of templates was inadequate and to fix the problem by adding more templates. The idea would be to build a set of templates that results in a relationship that parallels that in Equation 18 but with a fundamental difference: now Ω_0 is proportional to $def^{1/2}$ rather than to *def*. In this case, σ_{obs} will not longer be a constant. It is not difficult to introduce templates that accomplish this. A second way to modify the model, which is the approach adopted here, is to keep the templates as they are and change the function that relates Ω_{obs} to Ω_0. Until now we assumed that these variables were proportional to each other, but this was just a simplifying assumption that doesn't necessarily represent the best choice.

Let us instead assume, as we did at the end of the last section, that for planar surfaces $\Omega_{obs} \propto \Omega_0{}^{1/2}$ or $\Omega_{obs} = k\Omega_0{}^{1/2}$, where k is a constant. This results in (again, using Equation 18 and $\sigma_{obs} = def/\Omega_{obs}$):

$$\sigma_{obs} = \frac{\sqrt{\sigma_0}}{k\sqrt{h(\tau, \tau_0)}}\sqrt{def}, \tag{19}$$

which is in excellent agreement with the psychophysical data of Domini and Caudek (1999) (see their Figure 15). Another, simpler, option is to assume a linear relationship: $\Omega_{obs} = k\Omega_0 + \zeta$, where both k and ζ are constants.[12] This results in:

$$\sigma_{obs} = \frac{1}{\frac{kh(\tau, \tau_0)}{\sigma_0} + \frac{\zeta}{def}}. \tag{20}$$

As shown in Figure 2b, within the range of values tested the two options given by Equations 19 and 20 result in a good fit to the psychophysical data of Domini and Caudek (1999), shown in Figure 2a.

Domini and Caudek (1999) also found that the perceived slant, σ_{obs}, varies with the simulated tilt, τ, of the planar surface (see their Figure 18, shown here in

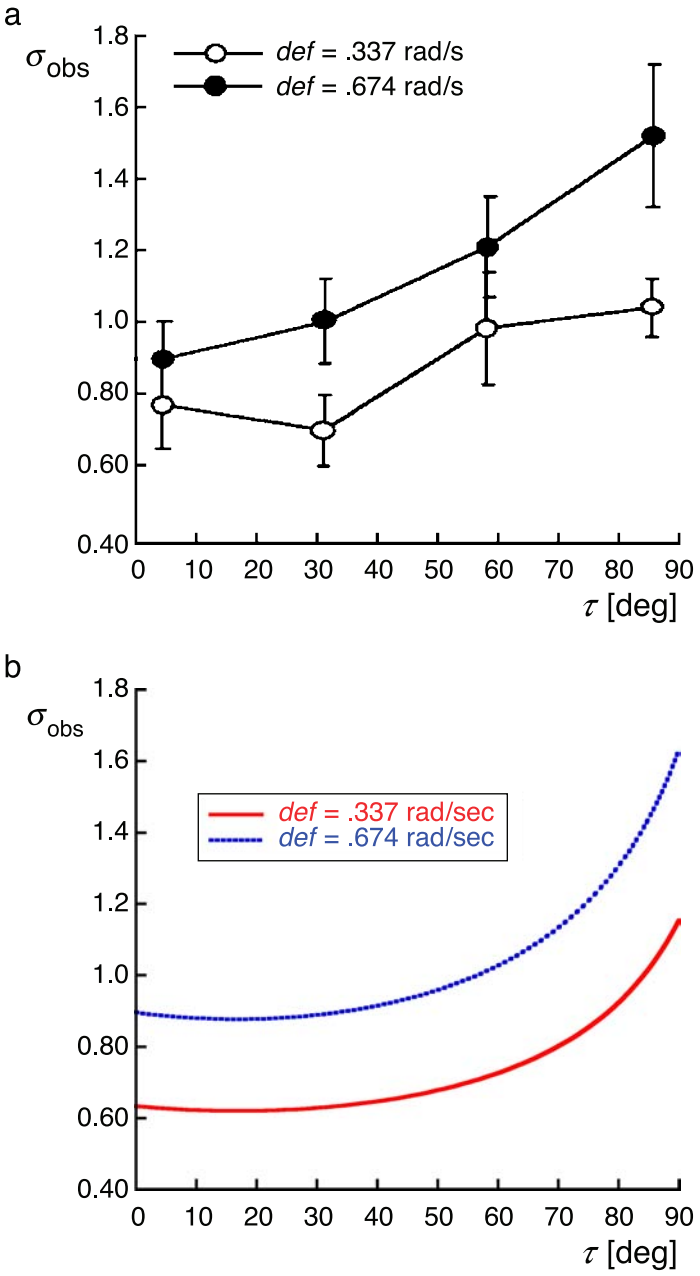

Figure 3. a) The perceived slant of a planar surface, σ_{obs}, as a function of the surface's simulated tilt, τ. Each curve corresponds to a different value of *def,* from Figure 18 of Domini and Caudek (1999). (Reproduced with permission). b) Model predictions of perceived slant, σ_{obs}, as a function of tilt, τ, for the two simulated values of *def* used in (a).

Figure 3a). Their heuristic is unable to explain this dependence but our model predicts it (see Equation 19 or 20). Figure 3b shows the predicted dependence of σ_{obs} on τ for Equation 19 (Equation 20 gives similar results). Figure 3b shows two curves, one for each of the two values of *def* used in Figure 18 of Domini and Caudek (1999). All templates are planar surfaces with $\tau_0 = 0$ and $\sigma_0 = \tan(75°)$. The match of predictions to the data is quite good, considering the simplifying assumptions used.

Perceived curvature of elliptical cylinders

Our implementation of the model includes a set of templates of elliptical cylinders of various ellipticities and angular speeds, all rotating about their longitudinal axis and oriented with the major axis of the ellipse making the same angle to the line of sight. For this case, Equation 15, which gives the averaged tuning parameter $\langle\Omega_0\rangle$, does not have an analytical solution and has to be solved numerically (see Notes on computational methods).

Figure 4a shows our recent psychophysical data (Fernandez & Farell, 2007) obtained from cylinders of different simulated ellipticities rotating about their longitudinal axes. The ratio C_{obs}/C_{sim} (where C_{sim} and C_{obs} are the simulated and the perceived curvature, respectively) is the depth-scaling factor, which was found to be a function of the stimulus' shape (i.e., of C_{sim}).[13]

We showed in the previous section that, for planar surfaces, our model predictions agree well with the model of Domini and Caudek (1999). This model states that the scaling factor is a decreasing function of *def*. Domini and Caudek's model works well for predicting the perceived slant of planes, but fails to explain the data for cylinders shown in Figure 4a. In such a case Domini and Caudek's model predicts that the scaling factor should decrease with C_{sim}; this is contrary to psychophysical results (Fernandez & Farell, 2007), which show an increase.

Figure 4b shows the template model predictions. The stimuli we used here consisted of cylinders of different ellipticities with their major axis making an angle of 12.5° with the line of sight, which corresponds to the average orientation of the rotating cylinders used to obtain the psychophysical data of Figure 4a.[14] Also assumed is a linear relationship between Ω_{obs} and Ω_0 (see Notes on computational methods). The quantitative match between data and model is very good, especially considering the simplifications made in choosing the family of templates.

Note that for the planar templates we assumed a square root relationship between Ω_{obs} and Ω_0. For cylinders, instead, we assumed a linear relationship. The two assumptions do not contradict each other, because the relationship can be different for different sets of templates: Different sets of templates are independent of each other. In the physiological implementation of the model that we will present in a following paper it will become apparent why this can be so. It just depends on how template wires its feedback connections to the lower areas, so the only constraints here are psychophysics and simplicity.

Perceived slant of the axis of rotation

Caudek and Domini (1998) measured the perceived slant of the axis of rotation. They found that, for planar surfaces, the slant of the perceived axis is a function of *def* (see their Figure 8, reproduced here in Figure 5a). Equation 6 gives the perceived slant of the axis of rotation

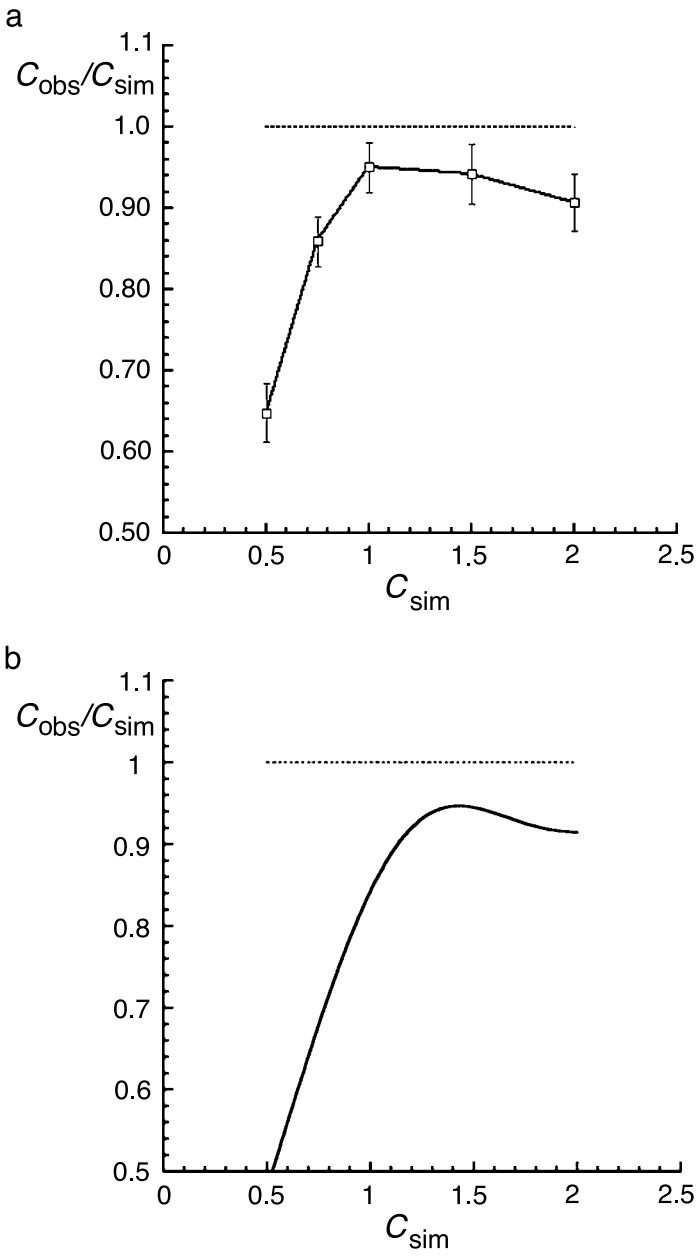

Figure 4. Perceived depth-scaling factor as a function of simulated curvature. a) Psychophysical data averaged across 4 subjects and across inclinations (the shape of the curve does not change with inclination). Redrawn from Fernandez and Farell (2007). b) Predicted data from our template-matching model, assuming that the perceived angular speed of rotation is proportional to the tuning parameter Ω_0.

as predicted by our model. In order to compare our predictions with the data of Caudek and Domini's (1998), we will define the slant as $\sigma_{axis} = 90 - \theta_{obs}$ (making σ_{axis} an angle, rather than the tangent of an angle, as it was in the previous sections). Then, Equation 6 becomes:

$$\cos\sigma_{axis} = \frac{-\partial_x v_y}{\lambda\Omega_{obs}}. \quad (21)$$

Caudek and Domini's (1998) used planar surfaces in which the axis of rotation is contained in the planar surface, so the normal to the surface was perpendicular to the axis of rotation. In such a case, we can extend the result $\Omega_{obs} = k\Omega_0^{1/2}$ (valid for the frontoparallel rotations) to the case in which the axis of

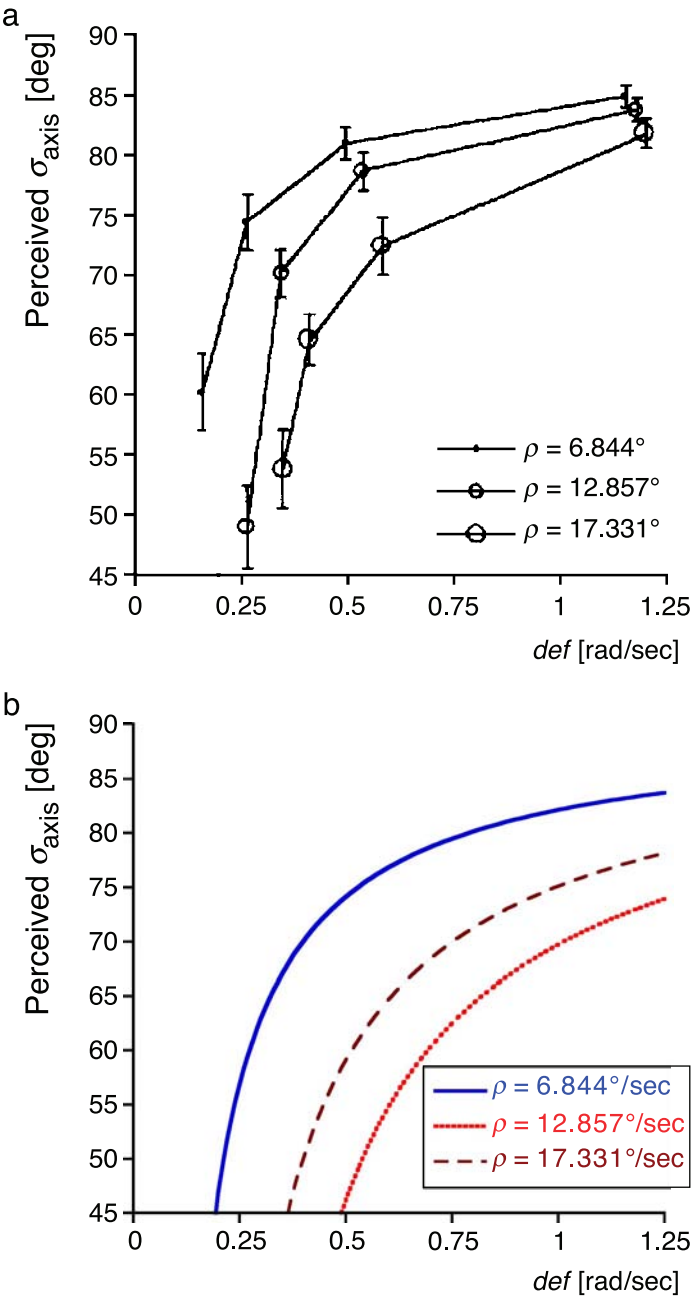

Figure 5. a) The perceived slant of the axis of rotation, σ_{axis} (measured from the line of sight), from Figure 8 of Caudek and Domini (1998) (reproduced with permission). The stimuli are planar surfaces, and perceived slant is plotted as a function of *def,* for the three values of ρ. b) Model predictions of the data shown in (a).

rotation is slanted. Then, it is easy to show that (see Appendix A):

$$\cos\sigma_{\text{axis}} = \frac{\rho}{\lambda C\sqrt{def}}, \quad (22)$$

where ρ (used as a parameter in Caudek & Domini, 1998) is the component of the speed of rotation along the line of sight (the z-component of Ω, which is assumed not to be observable), and C is a constant given that the stimulus tilt, τ, is constant ($C = k[h(\tau, \tau_0)/\sigma_0]^{1/2}$, see Equation 18).[15]

Figure 5b shows the Equation 22 predictions for the values of *def* and ρ used in Figure 8 of Caudek and Domini (1998), shown here as Figure 5a. The three curves were fitted using the same value of λC. There is good agreement between the psychophysical data and model predictions.

The model also predicts data for rotating cylinders. We have seen that, for cylinders, a good approximation is $\Omega_{\text{obs}} = c\Omega_0 = k\Omega_{\text{sim}}$ (see Equation 15). Using Equations 3 and 6 now becomes:

$$\sin\theta_{\text{obs}} = \frac{\sin\theta_{\text{sim}}}{\lambda k}, \quad (23)$$

where θ_{sim} is the simulated slant of the axis of rotation measured as a deviation from vertical. Figure 6 shows the predictions of Equation 23 together with the psychophysical data for circular cylinders obtained from three observers (data replotted from Fernandez & Farell, 2007). Each observer was fitted to a different value of λk. Again, we find a reasonably good match between psychophysical data and model predictions.

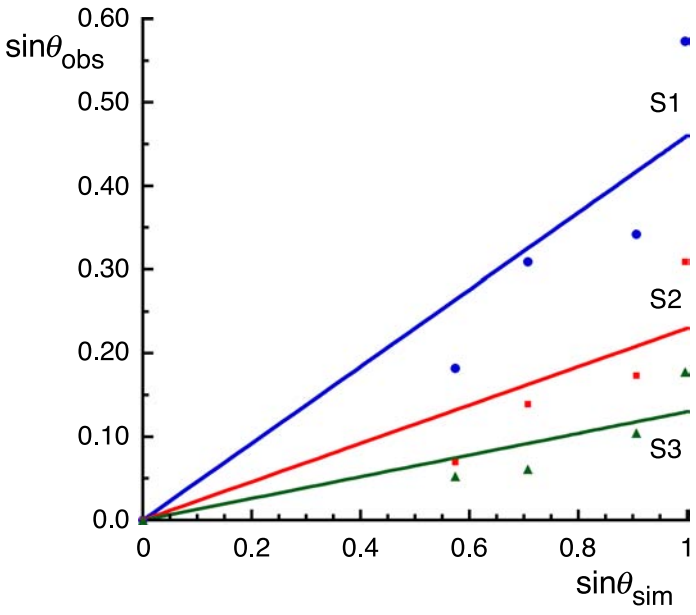

Figure 6. Model predictions (lines) from Equation 23, for circular cylinders. Predictions were fitted to the data of three subjects (triangles, circles and squares) by a least squares method. Psychophysical data were redrawn from Fernandez and Farell (2007).

We mentioned before that, for cylinders, the template matching model is consistent with the hypothesis that λ = constant. only if the condition $\sin\theta_{\text{obs}} \propto \sin\theta_{\text{sim}}$ is satisfied. Equation 23 satisfies this condition, showing that the template matching model is consistent with the hypothesis that λ = constant.

Discussion

Segmentation in SFM

In the Domini et al. (Domini & Braunstein, 1998; Domini & Caudek, 2003; Domini et al., 1998) heuristic perceived slant for planar surfaces is a function of *def*. There are two ways to extend this to curved surfaces. For curved surfaces, *def* is a function of retinal position (x, y), and changes from point to point. One extension—the continuous or *differential* version—is to assume that both the perceived slant and the perceived angular speed of rotation, Ω_{obs}, change locally from point to point as a function of local *def*. This is extremely important because the depth-scaling factor is proportional to the inverse of this angular speed. Because Ω_{obs} is computed locally, the structure of the recovered object is internally consistent but not affine. That is, the recovered object is not related to the simulated object by a simple scaling factor in depth. Rather, different scaling factors are found at different points on the recovered object. Note also that because the rotation rate is allowed to change from position to position, the object should be perceived as nonrigid.

A second extension—the discrete or *non-differential* version—assumes that the angular speed of rotation is computed as an average across differentiable surfaces, rather than locally. The logic here is that some objects might be perceptually segmented, with each patch possessing a different angular speed. Regardless of how the speed of rotation of a given surface patch is estimated, it is clear that the recovered structure will be affine for each patch, and each patch will be perceived as moving with its own angular speed, making the object non-rigid in the general case.

In Fernandez and Farell (2007), we showed that our psychophysical data for rotating cylinders rule out both extensions of the Domini et al. (Domini & Braunstein, 1998; Domini & Caudek, 2003; Domini et al., 1998) heuristic to curved surfaces. The data are inconsistent with perceived slant of such surfaces being computed (either locally or as an average over segmented patches) as a function of *def*. The template-matching model that we developed overcomes this problem by providing a unified framework for both planar and curved surfaces. For planar

surfaces the model predicts the dependence of perceived slant and Ω_{obs} on *def*, and for cylinders it reproduces the variation of perceived versus simulated curvature described in Fernandez and Farell (2007).

The model developed in this article corresponds to the discrete version. Here, before being fed into the model, the image is segmented by an algorithm based on segmentation psychophysics.[16] The depth structure of each segmented part is then computed independently of all others, using a single computed value of Ω_{obs} for each patch (though Ω_{obs} can differ across patches).

Support for the discrete version comes from the work of Di Luca, Domini, and Caudek (2004). They found that linking spatially separated patches into global entities affects the perception of local surface orientation. This result supports the discrete version over the continuous version of our model, for the latter hold that perceived surface orientation depends only on the local, rather than the global, optic-flow field.

Frontoparallel rotations: Current theories

As mentioned in the Introduction, a number of studies show that, for frontoparallel rotations, there is an affine relationship between the simulated and the perceived shape in SFM. The theory developed in this article agrees with this result and adds quantitative predictions lacking from previous theoretical treatments. We also mentioned in the Introduction the alternate conclusions reached in some studies (Domini & Braunstein, 1998; Domini & Caudek, 2003; Domini et al., 1998). These studies suggested that the perceived slant of a planar surface is a monotonically increasing sublinear function of *def*. As a result, the perceived depth separation between two points with a fixed physical depth separation on a planar surface would be a decreasing function of *def* (see Appendix A).

This claim generates a radical prediction: The recovered object's depth structure will be internally inconsistent. That is, the integral of the signed depths across a closed path will not sum to zero. Also predicted are distortions of depth-order relations and of parallelism, so that the recovered structure will be neither affine nor Euclidean. In a series of experiments, Domini and Braunstein (1998) and Domini et al. (1998) [referred together as Domini et al. in what follows] presented evidence to support these predictions. In a previous article (Fernandez & Farell, 2007) we showed that all the experiments preformed by Domini et al. could be reinterpreted under the hypothesis that in SFM the angular velocity is misperceived as varying between surfaces having a common axis of rotation. In that article we erroneously claimed that Domini et al. model assumed a single angular speed of rotation for the whole object. This mistake was motivated by the fact that Domini et al. assumed that the edge of an "open book" (two intersecting planar surfaces) will be perceived as just that, an edge. However, a difference in the perceived speed of rotation between surfaces predicts that, when the open-book edge is not contained in the same frontoparallel plane as the axis of rotation, then a gap in depth ought to be perceived between book's 'pages' where they are simulated to intersect. Domini et al.'s prediction of an inconsistent depth structure resulted from neglecting this gap. As we show in the Appendix A, the assumption that perceived slant is a function of *def* results in an internally consistent perceived structure once we take the gap into account. All the psychophysical data used by Domini et al. as a support for their internal-inconsistency model are also easily reinterpreted in terms of an internally consistent perceived depth structure.

There is an additional, related, issue that we need to address here. Orthographic projection is inherently ambiguous in an important but often neglected respect: the optic flow of a pure rotation is indistinguishable from that of a translation plus a rotation. Any pure rotational motion can be decomposed into a translation plus a rotation. This is so because we can subtract any instantaneous speed, identified as an overall translation, from the entire velocity field, and then interpret the modified velocity field that remains as a pure rotation. Each combination corresponds to a different depth structure. Suppose, for instance, that we have an open book that rotates around a frontoparallel axis. Suppose further that the axis of rotation is parallel to the edge at which the two planar surfaces meet, but the edge and the axis of rotation are not in the same frontoparallel plane. The edge, then, moves in a circular path around the actual axis of rotation. If we subtract the instantaneous speed of the edge, the resulting optic-flow field is consistent with an open book rotating around the edge. If the visual system decomposes the velocity field in this way, then our modification of the Domini et al. heuristic predicts that the edge would be perceived as such, rather than as a gap in depth. Thus, a pure-rotation interpretation of the optic-flow field of an open book predicts that a gap must be perceived at the edge, but an alternative rotation-plus-translation interpretation can always be found such that no gap is perceived.

We performed various informal experiments (not shown) whose results are consistent with the hypothesis that the visual system actually subtracts a constant speed to the optic-flow field before processing SFM. In one of them, for instance, two rigidly rotating planar surfaces simulated as having a gap in depth are perceived as joining at an edge. This happens if the motion of the surfaces is such that the retinal speeds of the two surfaces are the same at the point where the projections of the surfaces meet. Thus, even if an actual gap was explicitly simulated in this case, we do not see it. (There is an alternative interpretation. One can assume—in principle—that both surfaces are moving with the same angular speed, even if this is not the case across time—in our simulations with gap we used two surfaces moving at different angular speeds. If the perceptual system were doing that, the

stimulus would be consistent with a no-gap interpretation. But we assumed, in agreement with psychophysical data, that the visual system perceived the two planar surfaces as having different angular speeds. The visual system perceives different angular speeds for a dihedral angle even when the two surfaces are simulated as having the same angular speed—perceptually the open book either opens or closes at it moves. Thus, under that assumption, the only way to have a non-gap interpretation is to subtract the speed of the moving edge. To us this interpretation seems more consistent with the psychophysical data than to assume a no-gap interpretation based on perceiving the two surfaces as having the same angular speed. However, the later explanation, although implausible, cannot be ruled out.)

The theory of Domini et al. is consistent with a no-gap edge only for the special case in which the edge and the axis of rotation are coincident. Subtracting the speed of the edge from the optic flow field transforms the general case in which the edge and the axis of rotation are not coincident into the special case in which they are coincident. Thus, this transformation predicts that no gap should be observed and that the internal depth structure is consistent.

Domini et al. did not subtract any speed from the original optic flow field. Thus, a consistent interpretation of the flow field when the edge and the axis of rotation are not coincident requires that a gap be perceived. Ignoring the perception of a gap leads to the prediction of an inconsistent internal depth structure.

Beyond first-order optic flow

Our model was based on the analysis of the first-order optic flow under orthographic projection. However, there is evidence that humans can use second-order optic flow information (that is, accelerations) and perspective information in computing the SFM. Eagle and Hogervorst (1999) found this to be true for discriminations of the dihedral angle of a hinged plane (open book). This open book rotated about a frontoparallel axis. Discrimination thresholds decreased with size under perspective projection. Under orthographic projection, instead, thresholds increased with size until reaching random performance levels. On another study but using the same stimuli, Hogervorst and Eagle (1998) found a pattern of misperceptions that depended on stimulus size and rotation angle. They showed that errors in the estimate of speeds and accelerations could explain the pattern of misperceptions. Hogervorst and Eagle (2000) confirmed this result and extended it to orthographic projections.

Our model did not include these effects, which are certainly worthy of investigation, both psychophysically and by modeling. Another avenue for further research, suggested by a reviewer, would be to look at the consequences of ignored optic-flow parameters. Parameters that the model does not use should result in predicted metameric combinations of objects and motions. Do humans also have difficulty discriminating these combinations?

Conclusions

This article presents a quantitative model of the perception of structure from motion. Its predictions match most psychophysical data for both frontoparallel and non-frontoparallel rotations.

The model's structure is conceptually simple, and consists of two parts. The first part is an equation (Equation 7) that describes how to compute the perceived depth structure as a function of two global parameters, Ω_{obs} and ∇_{obs}. The second part is a template-matching model that uses the optic flow field as the input to compute the two global parameters required by Equation 7.

The model is algorithmic. Therefore, even though some of its computational modules, such as the MST templates, are inspired in physiology, the model is not intended as a physiologically plausible model of SFM computation. To fill this gap we have implemented a physiologically plausible neural model of SFM computation based on the model presented here; it will appear as a separate paper.

The model predictions presented here were derived from severely constrained assumptions about the templates. The general success of these predictions shows that the template approach is well suited for modeling the problem, even without an optimized template set. As we already mentioned, the addition of more templates should improve the fit of the model's predictions to psychophysical data.

Of course, the idea is not to have an infinite number of templates, but only a basic set. Visual objects will give rise to a population response across the set. For purposes of illustration, in this article we choose a minimal basis set, planes and cylinders, representatives of first-order and second-order templates, respectively. This is the simplest set consistent with our purposes. Any stimulus that is neither a cylinder nor a planar surface will fall into a separate parameterized class that is not represented in the templates. Nevertheless, each template in each set of templates (that is, templates for both cylinders and planar surfaces) will vote individually on which angular speed corresponds to the presented stimulus. In addition, we are currently trying a different set of templates resembling optic-flow operators, rather than specific objects. This approach would be more in line with the tuning properties of some MST cells. Beyond this, it would be rather premature to speculate about what might constitute a realistic set of basic templates.

Notes on computational methods

Computation of Figure 3b

We used Equation 19, where $h(\tau, \tau_0)$ is derived in the Appendix A. For the template we used a plane with $\tau_0 = 0$ and $\sigma_0 = \tan 75°$. The limits of integration used were $x_1 = y_1 = -.45x_2 = -.45y_2$ (where the particular value for x_1 is not important because they cancel out).

Computation of Figure 4b

The stimuli were cylinders with simulated curvatures, C_{sim}, with a range of [0.5, 2] and an angle of 12.5° between the line of sight and cylinder's apex (equal to the average angle used for the data of Figure 4a) and Ω_{sim} = 135°/sec (the value used in Fernandez & Farell, 2007).

For the templates, we approximated Ω_0 as being a continuous variable and used Equation 14 to compute Ω_0 for each of the curvatures used. We used cylinders with curvatures C = .5, 1, 1.5, 2, 2.5 and 3; and with an angle of 5° between the line of sight and cylinder's apex. We integrated numerically the second and third integrals of Equation 13 in order to compute Ω_0 in Equation 14. Then we used the activities computed from Equation 9 to compute the population average value for Ω_0, with $\beta = 5 \times 10^{-4}$ sec^2/deg^2.

Once Ω_0 was computed, we obtained Ω_{obs} as $\Omega_{obs} = \Omega_0$ (notice that $C_{obs}/C_{sim} = \Omega_{sim}/\Omega_{obs}$, see Fernandez & Farell, 2007).

Computation of Figure 5a

We used Equation 22 and fitted by hand the value of 0.2 for λC (same value for the three curves) to get a reasonable match with the psychophysical data shown in Figure 5b.

Computation of Figure 6

The data from Fernandez and Farell (2007) was replotted here and fitted by the least squares method to a straight line constrained to pass through the origin of coordinates.

Appendix A

Derivation of Equations 6, 18 and 22

Equation 6

By definition (Fernandez & Farell, 2007), $1/\lambda$ is a factor that represents the deviation between the perceived angular speed of rotation, Ω_{obs}, and an optic-flow consistent angular speed of rotation, Ω_{OFC} ($=-\partial_x v_y / \sin\theta_{obs}$). In general, Ω_{obs} is not consistent with the first-order optic flow. Substituting the definition of Ω_{OFC} into the definition of λ ($=\Omega_{OFC}/\Omega_{obs}$), we obtain Equation 6.

Equation 18

To compute Ω_0 we use Equation 14. To do so, we need to compute first the integrals A_G and B, defined in Equation 13. The distance of a point on a planar surface from a frontoparallel plane in a direction towards the observer can be written as (see, e.g., Domini & Braunstein, 1998):

$$z(x,y) = \frac{\sigma}{\sqrt{1+\tau^2}}(x+\tau y) = ax + by, \tag{A1}$$

where σ and τ are the slant and tilt of the planar surface. We will use Equation A1 to represent the stimuli, and a similar equation but with σ and τ replaced by σ_0 and τ_0 for the planar surface representing the template. Then we have

$$\begin{aligned} A_G &= \int_{x1}^{x2}\int_{y1}^{y2} (a_0x + b_0y)^2 dx\,dy \text{ and} \\ B &= \int_{x1}^{x2}\int_{y1}^{y2} (ax + by)(a_0x + b_0y) dx\,dy. \end{aligned} \tag{A2}$$

After lengthy but straightforward work integrating these two expressions, we get:

$$A_G = \sigma_0^2 f_A(\tau_0) \text{ and } B = \sigma\sigma_0 g_B(\tau, \tau_0), \tag{A3}$$

where

$$f_A(\tau_0) = \frac{\alpha_1}{1+\tau_0^2} + \frac{\alpha_2}{1+\frac{1}{\tau_0^2}} + \frac{\tau_0\alpha_3}{1+\tau_0^2} \text{ and}$$

$$g_B(\tau,\tau_0) = \frac{\alpha_1}{\sqrt{(1+\tau^2)(1+\tau_0^2)}} + \frac{\alpha_2}{\sqrt{\left(1+\frac{1}{\tau^2}\right)\left(1+\frac{1}{\tau_0^2}\right)}} + \frac{(\tau+\tau_0)\alpha_3}{\sqrt{(1+\tau^2)(1+\tau_0^2)}}, \tag{A4}$$

with

$$\begin{aligned} \alpha_1 &= \frac{1}{3}(x_2^3 - x_1^3)(y_2 - y_1) \\ \alpha_2 &= \frac{1}{3}(y_2^3 - y_1^3)(x_2 - x_1) \\ \alpha_3 &= \frac{1}{4}(x_2^2 - x_1^2)(y_2^2 - y_1^2). \end{aligned} \tag{A5}$$

Replacing Equations A3 and A4 into Equation 14 we get Equation 18, where $h(\tau, \tau_0) = g_B(\tau, \tau_0)/f_A(\tau_0)$.

Equation 22

By replacing $\theta_{obs} = 90 - \sigma_{axis}$ and $-\partial_x v_y = \Omega_{sim}\cos\sigma_{axis} = \rho$ in Equation 6, we obtain:

$$\cos\sigma_{axis} = \frac{\rho}{\lambda\Omega_{obs}}. \quad (A6)$$

Now, $\Omega_{obs} = k\Omega_0^{1/2}$, and from Equation 18 we have $\Omega_0 \propto def$, so $\Omega_{obs} = C\sqrt{def}$, which substituted into Equation A6 gives Equation 22.

Miscellaneous derivations

Condition required for having λ = const. (*cylinders*)

As shown in Fernandez and Farell (2007), for cylinders rotating about its longitudinal axis, the shape of the perceived object relative to the axis of rotation (Δz_{axis}) does not change with the inclination of this axis (Δz_{axis} = const.). As shown there also, this perceived structure is given by:

$$\frac{\Delta z_{axis}}{Z_0} = -\lambda\frac{\Delta v_x}{\partial_x v_y}\tan\theta_{obs}\cos\theta_{obs}, \quad (A7)$$

which is basically the first term in Equation 5 multiplied by $\cos\theta_{obs}$ to project it in the direction perpendicular to the axis of rotation. Replacing $v_y = v_z^{axis}\sin\theta_{sim}$ (see Fernandez & Farell, 2007), where v_z^{axis} does not change with a change in the axis's slant, Equation A7 becomes:

$$\lambda\frac{\Delta v_x}{\partial_x v_z^{axis}}\frac{\sin\theta_{obs}}{\sin\theta_{sim}} = \text{const.} \quad (A8)$$

Because $\Delta v_x/\partial_x v_z^{axis}$ = const. (that is, it does not change with the axis's slant, as long as the angular speed of rotation is kept constant), the condition for λ to be a constant becomes $\sin\theta_{obs}/\sin\theta_{sim}$ = const. When we allow not just the axis's inclination, but also the speed of rotation to change, then the condition for λ to be a constant becomes more complex.

The optic-flow field of a rotating object can be written as the product of a shape factor and the angular speed of rotation

This decomposition of optic flow is valid only for frontoparallel rotations. Let us assume, without loss of generality that the axis of rotation is vertical. Then, the trajectory of an arbitrary point on the object will follow a circular path (the circle, of radius R, is perpendicular to the axis of rotation). The coordinates of such a point can be described as:

$$\begin{aligned} x &= R\sin(\Omega t + \varphi) \\ y &= \text{const.} \\ z &= R\cos(\Omega t + \varphi), \end{aligned} \quad (A9)$$

where Ω is the angular speed of rotation and φ is the initial phase. The horizontal speed of such a point will be given by

$$v_x = \frac{dx}{dt} = \Omega R\cos(\Omega t + \varphi) = \Omega z(x, y), \quad (A10)$$

where we give z as a function of position (x, y) because the point under consideration is arbitrary and the description is valid for any point on the object. Equation A10 shows that the velocity field of a rotating object can be written as the product of a shape factor and the angular speed of rotation.

When the axis of rotation is slanted, v_x is still proportional to Ω, but in this case instead of a shape factor we have a more complex expression. This is easily obtained from Equation 2, which after a little work can be rewritten as

$$v_x = \Omega\left[\frac{z(x, y)}{Z_0}\cos\theta - y\sin\theta\right] = \Omega\Gamma(x, y, \theta), \quad (A11)$$

$def = \Omega\sigma = \Omega_{obs}\sigma_{obs}$

Domini and Braunstein (1998) showed that $def = \omega\sigma$ for arbitrary rotations (where ω is the projection of Ω into the frontoparallel plane). Thus this reduces to $def = \Omega\sigma$ for frontoparallel rotations. Here we show that $\Omega_{obs}\sigma_{obs} = def$ for frontoparallel rotations, which is the context in which we used it.

By definition,

$$def = \sqrt{def_1^2 + def_2^2}, \quad (A12)$$

where $def_1 = \frac{\partial v_x}{\partial x} - \frac{\partial v_y}{\partial y}$ and $def_2 = \frac{\partial v_x}{\partial y} + \frac{\partial v_y}{\partial x}$. For frontoparallel rotations $v_y = 0$, so def reduces to

$$def = \sqrt{\left(\frac{\partial v_x}{\partial x}\right)^2 + \left(\frac{\partial v_x}{\partial y}\right)^2} = |\vec{\nabla} v_x|, \quad (A13)$$

where $\vec{\nabla} v_x$ is the gradient of v_x. From Equation 1 we have

$$\Delta v_x = -\Omega_{obs}\frac{\Delta Z_{obs}}{Z_{obs}}, \quad (A14)$$

which shows that, at each location on the image, a change in v_x is maximal along the same direction (in the image plane) in which the change in perceived depth is maximal—or, stated in a different way, at each location in the image the gradients of v_x and z always point in the same direction, although this direction could vary across the image. If we call this direction r, then we have (from Equation A14):

$$\frac{\Delta v_x}{\Delta r} = \Omega_{\text{obs}} \frac{\Delta Z_{\text{obs}}}{\Delta r Z_{\text{obs}}}. \tag{A15}$$

By definition $\frac{\Delta v_x}{\Delta r} = \left| \vec{\nabla} v_x \right| = def$ and $\frac{\Delta Z_{\text{obs}}}{\Delta r Z_{\text{obs}}} = \sigma_{\text{obs}}$, which, when substituted into Equation A15, results in $def = \Omega_{\text{obs}} \sigma_{\text{obs}}$.

Perceived depth in planar surfaces is a decreasing function of def

Increasing slant increases perceived depth, but this happens if you keep a constant retinal separation between the points, which means that the simulated depth also increases. Thus, in this case perceived depth increases with simulated depth, as well. However, for fixed simulated depth but variable simulated slant (and thus variable simulated retinal size), the perceived depth is actually a decreasing function of simulated slant (and thus of *def*). Let us show this. By definition, we have:

$$\begin{aligned} \Delta z &= \sigma \Delta y \quad \text{and} \\ \Delta z_{\text{obs}} &= \sigma_{\text{obs}} \Delta y \end{aligned}. \tag{A16}$$

where σ is the surface slant, Δz the depth difference between two points on the surface and Δy the difference in height between the same two points (we assume here for simplicity that the surface has a tilt of zero; in the general case y would be measured along the tilt direction rather than along the vertical direction). The subscript *obs* refers to perceived quantities. If we eliminate Δy from Equation A16 and also use $def = \Omega\sigma$ and $\sigma_{\text{obs}} = \alpha\sigma^{1/2}$, where α is a constant, we get:

$$\Delta z_{\text{obs}} = \frac{\alpha \Omega^{1/2} \Delta z}{def^{1/2}}, \tag{A17}$$

which is a decreasing function of *def*.

Slant from *def* predicts internally consistent structure

The heuristic proposed by Domini et al. (1998) states that the perceived slant of a planar surface is a function of *def*. We show here that this predicts that the depth structure of the perceived object will be internally consistent.

Domini et al. (1998) heuristic is well defined for planar surfaces, or combinations of these, such as dihedral angles (open books). In this case, *def* has a constant value across each planar surface and so its value is well defined for the surface. The same is valid for the planar surface's slant, whose value is constant across the surface.

We will call this the discrete version of Domini et al. (1998), as opposed to the continuous version that can be used on curved surfaces. In the continuous version the value of *def* varies across the surface. It is unclear how the Domini et al.'s heuristic generalizes to this case. One way is for the perceived slant at any point on the surface to be a function of the local *def* at that position. Let us first deal with the discrete version of the model.

The discrete version of Domini et al.'s (1998) heuristic is actually similar to the one presented in this article: our model also predicts, in certain circumstances, that the slant is also a function of *def* (see Equations 19 and 20). Our model always results in an internally consistent perceived depth structure. Why then did Domini et al. (1998) claim that depth structure is internally inconsistent? Their claim is based in the assumption that the intersection (or edge) of two planar surfaces that are simulated to rigidly rotate with the same angular speed will be perceived to intersect at the simulated intersection even if they are perceived as rotating at different angular speeds. However, this is inconsistent with the basic SFM equations (see Fernandez & Farell, 2007).

We now show that the continuous version of the model also predicts an internally consistent depth structure. The easiest way to show this is the following. We need to show that the depth difference between any two points on the object is independent of the path that we use to compute this distance. The depth between two points can calculated as the path integral:

$$\Delta z_{12} = \int_{\text{P1}}^{\text{P2}} \text{d}z = \int_{\text{P1}}^{\text{P2}} \left(\frac{\partial z(x,y)}{\partial x} \text{d}x + \frac{\partial z(x,y)}{\partial y} \text{d}y \right), \tag{A18}$$

where $z(x, y)$ is an arbitrary function of retinal position (x, y). For instance, z could be a function of either *def* or σ, which in turn could also be arbitrary functions of retinal position (x, y). In general, the value of a path integral of the form $\int_{\text{P1}}^{\text{P2}} [\text{f}(x, y)\text{d}x + \text{g}(x, y)\text{d}y]$, with $\text{f}(x, y)$ and $\text{g}(x, y)$ arbitrary functions, is a function of the path from P_1 to P_2. To have a consistent object, the value of the integral in Equation A18 must be independent of the path. Using Stokes' theorem (Kaplan, 1952), it can be shown that a path integral is independent of the path if and only if

$$\frac{\partial \text{f}}{\partial y} = \frac{\partial \text{g}}{\partial x}, \tag{A19}$$

and for the case of Equation A18 this reduces to

$$\frac{\partial}{\partial y}\left(\frac{\partial z}{\partial x}\right) = \frac{\partial}{\partial x}\left(\frac{\partial z}{\partial y}\right). \quad \text{(A20)}$$

Equation A20 is always true if $z(x, y)$ is a continuous function of both x and y. Thus, we have demonstrated that, for a smooth surface, the perceived internal depth structure for the generalized Domini et al. model is always consistent.

For planar surfaces $\sin\theta_{obs} \propto \sin\theta_{sim}$

This is valid only for planar surfaces in which the axis of rotation is contained in the planar surface. For such a case, we found that a good match to psychophysical data is obtained if $\Omega_{obs} = k\Omega_0^{1/2}$. Following a procedure similar to the one used to obtain Equation 23, we get:

$$\sin\theta_{obs} = \frac{\Omega_{sim}^{1/2}}{\lambda\, k}\sin\theta_{sim}. \quad \text{(A21)}$$

Note that $\sin\theta_{obs} = \cos\sigma_{axis}$, thus Equation A21 is just a different version of Equation 22.

Acknowledgments

This research was supported by NEI Grant EY12286 (B.F.).

Commercial relationships: none.
Corresponding author: Julian M. Fernandez.
Email: julian_fernandez@isr.syr.edu.
Address: Institute for Sensory Research, Syracuse University, 621 Skytop Rd, Syracuse, NY 13224, USA.

Footnotes

[1]We use the term "affine" here in the more restricted way often used in SFM studies, which means a scaling factor in depth. An actual affine transformation can include a shearing transformation in depth, which does not preserve depth ordering.

[2]In principle, observers could use information about the period to estimate Ω. They could do this if they were able to see a few full rotations (which is not the case in most psychophysical tests) and the object had a salient feature to mark the periodicity. Our pilot experiments with rotating elliptical cylinders show that observers are very poor at predicting when a uniquely colored dot will reappear on the front. Even rather large shifts in the occluded dot's position go undetected when the dot reappears, making it unlikely that Ω is estimated from the period.

[3]This argument could explain why Hogervorst et al. (1997) found that stimuli in which the slant of the rotation axis was changing continuously did not lead to a nonrigid percept, in opposition to the results of Loomis and Eby (1988) and to the predictions of our model.

[4]An alternative way to compare Equations 4 and 5 is to say that the perceived 3-D structure and motion are consistent with a different optic flow in which the observed value of $(\Delta v_x/\partial_x v_y)$ is λ times the real value. This would mean that Δv_x, $\partial_x v_y$ or both are observed wrongly, whereas our approach assumes the observed $\partial_x v_y$ is $1/\lambda$ times the real $\partial_x v_y$.

[5]By "cylinder" we mean any shape that doesn't change along the axis of rotation. This includes planar surfaces in which the axis of rotation is contained in the planar surface.

[6]Of course, our theory is general and does not assume that the axis of rotation is fixed relative to the object as the slant of the axis changes. We only made the assumption that the axis of rotation is parallel to the longitudinal axis for cylinders because the psychophysical data from Fernandez and Farell (2007) were obtained under this condition. The model can also be applied to cylinders in which the longitudinal and the rotation axes are not linked. It will result in a different set of predictions, but at this time there are no psychophysical data to test them.

[7]Note that, in general, F and G are not identical functions and thus the set of parameters α and γ_0 need not be the same.

[8]The "curve" is actually an N-dimensional hypersurface in a N + 1 dimensional space, where N in the number of parameters γ_0.

[9]Strictly speaking, $\langle\Omega_0\rangle$ should be computed as the standard weighted population average $\langle\Omega_0\rangle = \frac{\int \Omega_0 r d\Omega_0 d\gamma_0}{\int r d\Omega_0 d\gamma_0}$; Equation 15 corresponds to an approximation in which only the peaks of the population response are considered, which is equivalent to treating the width of the population response along the two parameters Ω_0 and γ_0 as zero.

[10]An anonymous reviewer noted an interesting parallel between the template model and the model of Hogervorst and Eagle (1998). To quote the reviewer, "When the templates reflect the prior probability of shapes and motions that can occur (which is likely) the activity of the template cells resembles the posterior probability. Moreover, the model by Hogervorst and Eagle (1998) calculates for each combination of shape and motion parameters a probability that depends on the match between the optic flow (the stimulus) and the flow resulting from a 3-D object and motion (similar to a template). In the template model, each combination is covered by a template cell."

[11]We are using here Equation 14 instead of Equation 15 because all the templates used for planar surfaces have the

same value of the parameters σ_0, and τ_0, so no averaging is needed.

[12]This linear relationship is only meant as an approximation within the range of speeds tested. It cannot be valid when Ω gets close to zero. Notice that Ω_{obs} is minimal at about 12 deg/s and $\zeta = 0.5$ deg/s. In addition, there aren't any templates that represent $\Omega_{obs} = 0$ (similarly, there probably aren't any MST cells that respond to static stimuli). And, of course, at sufficiently low speeds SFM no longer works (neither perceptually nor in the model).

[13]C_{sim} in our experiments is a function of the curvature, but it is actually defined as the ratio between the two principal axes: the one that crosses the line of sight and the one that crosses the frontoparallel plane. The later has a constant size, so with this definition the ratio C_{obs}/C_{sim} gives the scaling factor in depth.

[14]The cylinders rotated about their longitudinal axes, which were frontoparallel (because, as explained elsewhere in the article, model results for cylinders are independent of the simulated slant of the axis of rotation). Thus, the elliptical cross-section of the cylinder (perpendicular to the longitudinal axis) is in the horizontal plane. The angle between the major axis of this ellipse and the line of sight is 12.5° regardless of the cylinder's ellipticity.

[15]As shown in Figure 2a, either a square root function or a linear function describes well the psychophysical data; we chose the square root function for analytical reasons, for the derivations result in simpler equations.

[16]Segmentation is based on the optic flow. However, the algorithm we used is very rudimentary. For instance, we segment into parts when there is a discontinuity in the first derivative of the optic flow, such as at the edge of an open book. But we are still far from a good understanding on how segmentation works in SFM (or how it works in general, for that matter).

References

Braunstein, M. L., Liter, J. C., & Tittle, J. S. (1993). Recovering three-dimensional shape from perspective translations and orthographic rotations. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 598–614. [PubMed]

Caudek, C., & Domini, F. (1998). Perceived orientation of axis of rotation in structure-from-motion. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 609–621. [PubMed]

Cornilleau-Pérès, V., & Droulez, J. (1989). Visual perception of curvature: Psychophysics of curvature detection induced by motion parallax. *Perception & Psychophysics, 46,* 351–364. [PubMed]

Di Luca, M., Domini, F., & Caudek, C. (2004). Spatial integration in structure from motion. *Vision Research, 44,* 3001–3013. [PubMed]

Domini, F., & Braunstein, M. L. (1998). Recovery of 3-D structure from motion is neither Euclidean nor affine. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1273–1295. [PubMed]

Domini, F., & Caudek, C. (1999). Misperceptions of angular velocities influence the perception of rigidity in the kinetic depth effect. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 426–444. [PubMed]

Domini, F., & Caudek, C. (2003). 3-D structure perceived from dynamic information: A new theory. *Trends Cognitive Science, 7,* 444–449. [PubMed]

Domini, F., Caudek, C., & Proffitt, D. R. (1997). Misperceptions of angular velocities influence the perception of rigidity in the kinetic depth effect. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 1111–1129. [PubMed]

Domini, F., Caudek, C., & Richman, S. (1998). Distortions of depth-order relations and parallelism in structure from motion. *Perception & Psychophysics, 60,* 1164–1174. [PubMed]

Eagle, R. A., & Blake, A. (1995). Two-dimensional constraints on three-dimensional structure from motion tasks. *Vision Research, 35,* 2927–2941. [PubMed]

Eagle, R. A., & Hogervorst, M. A. (1999). The role of perspective information in the recovery of 3D structure-from-motion. *Vision Research, 39,* 1713–1722. [PubMed]

Fernandez, J. M., & Farell, B. (2007). Shape constancy and depth-order violations in structure from motion: A look at non-frontoparallel axes of rotation. *Journal of Vision, 7*(7):3, 1–18, http://journalofvision.org/7/7/3/, doi:10.1167/7.7.3. [PubMed] [Article]

Fernandez, J. M., Watson, B., & Qian, N. (2002). Computing relief structure from motion with a distributed velocity and disparity representation. *Vision Research, 42,* 883–898. [PubMed]

Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements and actions. *Nature Reviews Neuroscience, 4,* 179–192. [PubMed]

Hildreth, E. C., Ando, H., Andersen, R. A., & Treue, S. (1995). Recovering three-dimensional structure from motion with surface reconstruction. *Vision Research, 35,* 117–137. [PubMed]

Hogervorst, M., Kappers, A. M. L., & Koenderink, J. J. (1993). Perception of metric depth from motion parallax. *Perception, 22,* 101.

Hogervorst, M. A., & Eagle, R. A. (1998). Biases in three-dimensional structure-from-motion arise from noise in the early visual system. *Proceedings: Biological Science, 265,* 1587–1593. [PubMed] [Article]

Hogervorst, M. A., & Eagle, R. A. (2000). The role of perspective effects and accelerations in perceived three-dimensional structure-from-motion. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 934–955. [PubMed]

Hogervorst, M. A., Kappers, A. M., & Koenderink, J. J. (1997). Monocular discrimination of rigidly and nonrigidly moving objects. *Perception & Psychophysics, 59,* 1266–79. [PubMed]

Kaplan, W. (1952). *Advanced calculus*. Reading, Massachusetts: Addison-Wesley Publishing Company, Inc.

Koenderink, J. J., & van Doorn, A. J. (1991). Affine structure from motion. *Journal of the Optical Society of America A, Optics and Image Science, 8,* 377–385. [PubMed]

Liter, J. C., Braunstein, M. L., & Hoffman, D. D. (1993). Inferring structure from motion in two-view and multi-view displays. *Perception, 22,* 1441–1465. [PubMed]

Loomis, J. M., & Eby, D. W. (1988). Perceiving structure from motion: Failure of shape constancy. In *Proceedings of Second International Conference on Computer Vision* (pp. 383–391). Washington, D.C.: Computer Society of the IEEE.

Norman, J. F., & Lappin, J. S. (1992). The detection of surfaces defined by optical motion. *Perception & Psychophysics, 51,* 386–396. [PubMed]

Norman, J. F., & Todd, J. T. (1993). The perceptual analysis of structure from motion for rotating objects undergoing affine stretching transformations. *Perception & Psychophysics, 53,* 279–291. [PubMed]

Perrone, J. A., & Stone, L. S. (1998). Emulating the visual receptive-field properties of MST neurons with a template model of heading estimation. *Journal of Neuroscience, 18,* 5958–5975. [PubMed] [Article]

Pollick, F. E., Nishida, S., Koike, Y., & Kawato, M. (1994). Perceived motion in structure from motion: Pointing responses to the axis of rotation. *Perception & Psychophysics, 56,* 91–109. [PubMed]

Simoncelli, E. P., & Heeger, D. J. (1998). A model of neuronal responses in visual area MT. *Vision Research, 38,* 743–761. [PubMed]

Tittle, J. S., Todd, J. T., Perotti, V. J., & Norman, F. N. (1995). Systematic distortion of perceived three-dimensional structure from motion and binocular stereopsis. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 663–678. [PubMed]

Todd, J. T. (1998). Theoretical and biological limitations on the visual perception of 3D structure from motion. In T. Watanabe (Ed.), *High-level motion processing—Computational, neurophysiological and psychophysical perspectives* (pp. 359–380). Cambridge, MA: MIT Press.

Todd, J. T., & Bressan, P. (1990). The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Perception & Psychophysics, 48,* 419–430. [PubMed]

Todd, J. T., & Norman, J. F. (1991). The visual perception of smoothly curved surfaces from minimal apparent motion sequences. *Perception & Psychophysics, 50,* 509–523. [PubMed]

Treue, S., Andersen, R. A., Ando, H., & Hildreth, E. C. (1995). Structure-from-motion: Perceptual evidence for surface interpolation. *Vision Research, 35,* 139–148. [PubMed]

Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.

Ullman, S. (1984). Maximizing rigidity: The incremental recovery of 3-d structure from rigid and nonrigid motion. *Perception, 13,* 255–274. [PubMed]

Werkhoven, P., & van Veen, H. A. (1995). Extraction of relief from visual motion. *Perception & Psychophysics, 57,* 645–656. [PubMed]