



Bootstrapping a better slant: A stratified process for recovering 3D metric slant

Xiaoye Michael Wang¹ · Mats Lind² · Geoffrey P. Bingham¹

Published online: 10 September 2019
© The Psychonomic Society, Inc. 2019

Abstract

Lind et al. (*Journal of Experimental Psychology: Human Perception and Performance*, 40 (1), 83, 2014) proposed a bootstrap process that used right angles on 3D relief structure, viewed over sufficiently large continuous perspective change, to recover the scaling factor for metric shape. Wang, Lind, and Bingham (*Journal of Experimental Psychology: Human Perception and Performance*, 44(10), 1508-1522, 2018) replicated these results in the case of 3D slant perception. However, subsequent work by the same authors (Wang et al., 2019) suggested that the original solution could be ineffective for 3D slant and presented an alternative that used two equidistant points (a portion of the original right angle). We now describe a three-step stratified process to recover 3D slant using this new solution. Starting with 2D inputs, we (1) used an existing structure-from-motion (SFM) algorithm to derive the object's 3D relief structure and (2) applied the bootstrap process to it to recover the unknown scaling factor, which (3) was then used to produce a slant estimate. We presented simulations of results from four previous experiments (Wang et al., 2018, 2019) to compare model and human performance. We showed that the stratified process has great predictive power, reproducing a surprising number of phenomena found in human experiments. The modeling results also confirmed arguments made in Wang et al. (2019) that an axis of mirror symmetry in an object allows observers to use the recovered scaling factor to produce an accurate slant estimate. Thus, poor estimates in the context of a lack of symmetry do not mean that the scaling factor has not been recovered, but merely that the direction of slant was ambiguous.

Keywords Bootstrap process · Geographical slant perception · Affine geometry · Stereomotion · Structure-from-motion · Symmetry

Introduction

Three-dimensional (3D) shape perception is an essential focus for investigations of human visually guided interactions with the environment. Many studies in 3D shape

Statement of Significance Empirically, the bootstrap process has been demonstrated to be an extremely replicable phenomenon through which human observers are able to perceive veridical 3D shape and slant. The current study offers a computational basis for this process, combined with simulation results from previous human experiments. Results from this study show how observers should behave when there is a need to perceive the veridical Euclidean structure of the surrounding environment.

✉ Xiaoye Michael Wang
wang492@indiana.edu

¹ Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA

² Department of Information Technology, Uppsala University, Uppsala, Sweden

perception have reported inaccuracy in judgment of 3D shape, where the perceived shape tends to be compressed or expanded in depth (e.g., Domini & Caudek, 2013; Koenderink & van Doorn, 1991; Todd & Bressan, 1990; Todd & Norman, 1991; Todd, Oomes, Koenderink, & Kappers, 2001; Wagner, 1985). This phenomenon prompted researchers to argue that the perceptual space should be affine or of relief.¹ However, if this were the case, daily tasks such as reaches-to-grasp objects would be seriously impaired, as they require accurate perception of metric shape to scale the grasp formed on approach to an object, which is most typically grasped back-to-front more than side-to-side. This was shown by

¹ The perceptual phenomenon is that distance is scaled by an unknown scaling factor along the depth dimension while being metrically specified in the projection plane. The conventional term used to describe this has been “affine” but more operations are allowed by the perceptual phenomenon than those defined by an affine mapping only. Wang, Lind, and Bingham (2018) adopted the word *relief* to better describe the phenomenon. See Appendix A in Wang et al. (2018) for an extended discussion of this issue.

Lee, Crabtree, Norman, and Bingham (2008) to be the case for feed-forward reaches-to-grasp, where the authors argued that poor shape perception resulted in poor reaches-to-grasp when the hand was not visible during the reach and became accurate when it was, allowing the use of online guidance. Ultimately, the fact that most studies report poor 3D shape perception is at odds with the common functionally effective and efficient performance of reaches-to-grasp, whether the hand is visible or not. Indeed, Lee and Bingham (2010) subsequently revealed conditions that yielded good levels of performance.

Bingham, Lind, and colleagues suggested that the discrepancy might be the result of experimental design because most studies that report such results presented their stimuli either statically or with a small amount of relative motion, while people in their daily activities are moving substantially relative to their surroundings (Lee & Bingham, 2010; Lee, Lind, Bingham, & Bingham, 2012; Lind et al., 2014). To address this issue, their initial idea was as follows: Because distance can be perceived accurately in a frontoparallel plane, but not in the depth direction, one can simply exchange depth and width by rotating the object, or moving the observer relative to the object, by 90° , so that the depth, now placed in a frontoparallel plane, can be perceived accurately. Upon investigation, this worked! They then tested a series of smaller amounts of relative motion and found that with continuous perspective change of approximately 45° or greater, people could perceive metric depth, and thus shape, accurately.

Lind et al. (2014) proposed a bootstrap process to account for this phenomenon, a process that is based the notion of *invariance over transformation*. This model first assumes that observers were able to recover 3D relief structure through two-frame apparent motion, which a number of existing structure-from-motion (SFM) models can achieve (e.g., Koenderink & van Doorn, 1991; Shapiro, Zisserman, & Brady, 1995; Lind, 1996). The resulting depth structure is scaled by an unknown factor compared to the actual depth. To recover this scaling factor, one first identifies two points on the object that are equidistant to the observer, an operation allowed in the relief frame of reference. With the assumption of small visual angle and, therefore, scaled orthographical projection, the line formed between these two points is orthogonal to the line of sight. Consequently, a third point can be identified on the object that is on a line parallel to the line of sight. These three points form a right angle that can be tracked across perspective change. Assuming rigidity, the angle formed among these three points would remain as 90° on the object but would be constantly changing in the relief space as a result of the interaction between motion in depth and the relief scaling factor. Subsequently, one can relate the new angle to the original 90° angle to recover the scaling factor. When the perspective change is small, the new angle typically

fails to deviate from the original angle by a sufficient amount to discriminate such change due to motion measurement noise. However, with sufficiently large perspective change, variations in the target angle in relief space would be large and, therefore, produce relatively more stable and accurate estimation of the scaling factor. In addition, the unknown scaling factor renders the amount of rotation as unknown, and consequently observers cannot know when perspective change is sufficiently large to “trust” the resulting scaling factor estimate. Therefore, there has to be something to inform them when this has occurred. Because angle bisection is allowed in relief space, the bisection of the target angle by the line of sight can be detected to inform observers that a 45° perspective change has occurred, and the recovered scaling factor should be accurate and reliable.

The experimental results consistent with the bootstrap process have been replicated in a number of different studies and tasks, including reaches-to-grasp wooden elliptical cylinders varying in depth-to-width aspect ratio (Lee & Bingham, 2010), recognition of complex computer-generated polyhedrons evaluated in terms of both errors and reaction times (Lee, Lind, Bingham, & Bingham, 2012), judgments of 3D depth-to-width aspect ratios of computer-generated 3D elliptical cylinders and asymmetric polyhedrons (where judgments were performed by adjusting the aspect ratio of an outline of the shape) (Lind et al., 2014), and, finally, a number of 3D slant judgment studies (Wang et al., 2018). Wang et al. (2018) replicated predictions of the bootstrap model using a 3D slant judgment task. Initially, they tested judgments of strictly planar surfaces, which failed to yield accurate judgments with large continuous perspective change. The authors reasoned that this was because SFM models require at least four non-coplanar points to obtain the relief structure that is necessary for the subsequent bootstrap process and planar surfaces lack non-coplanar points by definition. So, they introduced additional 3D non-coplanar structure to the planar surface, and the behavioral results were then just as the bootstrap process would predict: participants did produce accurate slant judgments with large continuous perspective changes approximately equal to or greater than 45° .

Again, what one needs to note is that the essence of the bootstrap process is not the 45° of continuous perspective change or bisecting an arbitrary right angle on the object; instead, it is identifying and consequently tracking an *invariant over the SFM transformation*. In the context of bootstrapping metric depth via relief depth, the invariant corresponds to identifiable metric or rigid physical structure of the object, while the transformation corresponds to the constant changes in the perceived relief structure as a result of relative motion in depth due to continuous perspective change. For the bootstrap process to work, one has to identify certain depth structures that are equivalent in both the relief and metric space at some instant of time. Such an equivalence offers a

bridge that links relief space to metric space. In the solution presented by Lind et al. (2014), the right angle acts as such a bridge. Because depth is scaled in relief space by the same scaling factor, albeit unknown, identifying two equidistant points is allowed. Also, assuming a small visual angle and therefore scaled orthographical projection, a third point can also be found such that the three points form a right angle. This angle is also of 90° in the physical space and therefore should remain invariant as the object moves relative to the observer. However, because relative motion creates depth changes, this angle would no longer be 90° in relief space. The observer, therefore, can use such transformation to determine the scaling factor that would yield the angle as 90° again.

One should also note that the principle behind the bootstrap process does not limit the method through which it is implemented. What is important is the process itself: *tracking an invariant and identifiable 3D structure in relief space through relative motion and using the resulting depth variations in relief space to extrapolate to the metric space*. The key is, therefore, to be able to obtain relief structure in the first place. The slant perception study by Wang et al. (2018) demonstrated this. They initially used strictly planar surfaces, slanted around the x-axis and rotated around the y-axis, to produce perspective change. This failed to yield an effect of large continuous perspective change given a lack of non-coplanar points with which to produce a relief depth map. Indeed, with the addition of 3D structure to the top of the planar surface, performance improved with large continuous perspective change showing that the bootstrap process can be applied to slant perception. Aside from successfully replicating previous findings, their results were also at odds with the traditional definition of slant.

Defining slant

Slant is studied in the context of the perception of surface orientation. As shown in Fig. 1, surface orientation can be decomposed into three components: slant (σ), tilt (τ), and roll (ω). Traditionally, slant has been defined as the angle formed between the line of sight, which corresponds to the z-axis, and surface normal (\vec{n}). Tilt is the angle between the projection of the surface normal on the image plane (\vec{n}_{xy}) and the x-axis, indicating the direction of slant (i.e., the direction of the maximum increase in distance relative to the line of sight) (Stevens, 1983a, 1983b). Roll is simply the angle of rotation of the surface around its normal. For a planar surface, roll is rather hard to define due to a lack of reference direction and is not as commonly studied as slant and tilt. Defining surface orientation this way can be found in many different studies, including those of 3D shape perception (e.g., Gibson, 1950; Gibson & Cornsweet, 1952; Stevens, 1983b; Saunders & Knill, 2001; Norman et al., 2006; Sawada & Pizlo, 2008). In

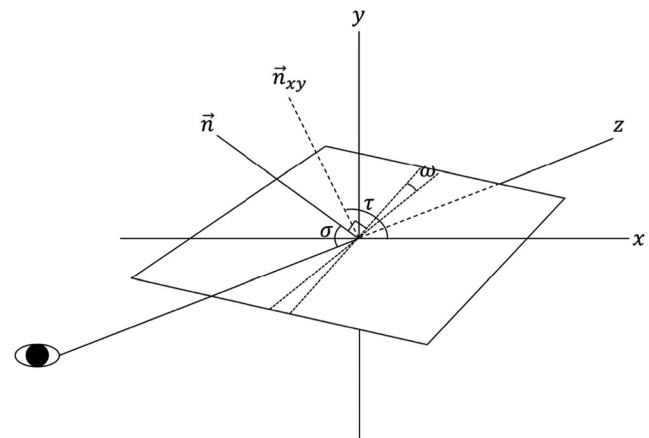


Fig. 1 Illustration of variables that are used to define surface orientation. Slant (σ) is defined as the angle formed between surface normal \vec{n} and the line of sight. Tilt (τ) is defined as the angle between the projection of the surface normal on the xy-plane \vec{n}_{xy} and the x-axis. Roll (ω) is the angle of rotation of the surface round its normal

fact, many have argued that slant defined this way should be considered as the basis for 3D shape perception (e.g., Koffka, 1935; Gibson, 1950; Beck & Gibson, 1955; Wallach & Moore, 1962; Kaiser, 1967; Hoffman & Richards, 1984; Todd, 2004; Sakata, Tsutsui, & Taira, 2005; Welchman, Deubelius, Conrad, Bültorf, & Kourtzi, 2005). We will refer to this definition of slant as *local slant*, since it is contingent upon the observer's line of sight. It has also been called *optical slant*. In this definition, slant is egocentric.

The issue with adopting a local frame of reference in the experimental paradigm that has been used to study the bootstrap process is that the local slant itself would not remain constant over the SFM rotation (or perspective changes). The objects used in Wang et al. (2018) contained planar surfaces that were first slanted around the x-axis and rotated around a vertical axis that was parallel to the y-axis and went through the center of the object. Assuming the observer fixates on the center of the surface, as the object rotates, the angle formed between the line of sight and surface normal would be constantly changing. Figure 2 shows the local slant of a surface slanted 30° at 90° local tilt (i.e., the face of the surface is facing the observer) as the surface rotates around a vertical axis that passes through its center. Throughout the rotation, the surface maintains a 30° angle to the ground plane. As can be seen from this figure, there is a large variation in local slant ($\approx 10^\circ$) as the surface rotates 45° (i.e., 45° change in tilt). Therefore, using a local frame of reference to describe slant in the context of SFM is not effective and could be extremely confusing simply because the slant, if defined locally, is not constant.

The angle that remains constant as the object rotates is the angle formed between the object and the ground plane (the xz-plane). Mathematically, this angle can be found through computing the angle between the object's surface normal and the

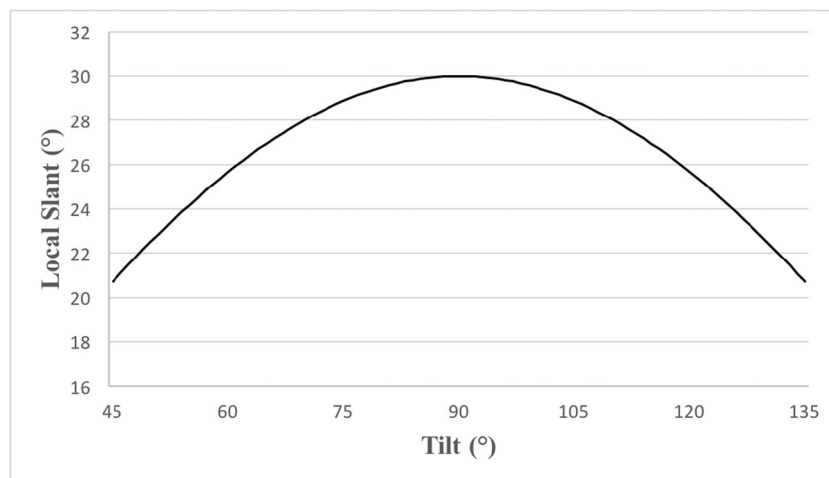


Fig. 2 Local slant variation as a surface that forms a 30° angle with the ground plane rotates around a vertical axis that passes through its center; 90° is when the surface is facing the observer’s line of right

ground plane’s normal, which can be described using the unit vector pointing in the y direction [0, 1, 0]. Correspondingly, tilt in this frame of reference should be defined as the projected surface normal on the xz-plane and the unit vector pointing toward the z direction [0, 0, 1]. Figure 3 shows this. We refer to these as *geographical slant* and *tilt* because slant remains invariant as the line of sight changes and tilt describes the direction of the gradient of slant relative to the ground surface. Based on this description, a geographical slant of 90° would be an upright surface and a slant of 0° would be horizontal (parallel to the ground). As the slanted surface rotates around a vertical rotational axis, which coincides with the y-axis, its geographical slant would remain constant while its local slant would be constantly changing.

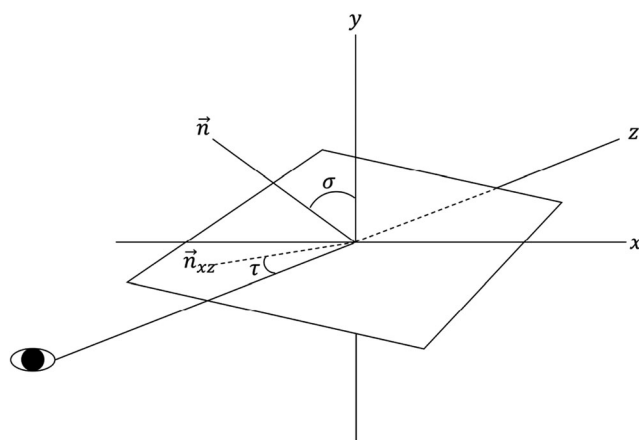


Fig. 3 Illustration of geographical variables in the current context. Geographical slant (σ) is defined as the angle formed between surface normal \vec{n} and the positive y-axis. Geographical tilt (τ) is defined as the angle between the projection of the surface normal on the xz-plane \vec{n}_{xz} and the z-axis

Issues with the right-angle solution lead to an alternative formulation

Even with an alternative definition of slant, Wang et al. (2019) noted that there was a problem inherent to the right-angle solution when it was applied to slant perception, which would render this solution relatively ineffective. Remember, the role of large continuous perspective change in the bootstrap process is to produce sufficiently large depth variations in the relief space on the invariant depth structure, for example, the right angle, that would subsequently allow perceivers to bootstrap to metric. This means that there needs to be sufficiently large depth variation in the right angle. This solution was originally proposed to account for performance in depth-to-width aspect ratio judgment experiments, where the objects were 3D polyhedrons with a horizontal top and the viewing angle, defined as the angle between the line of sight and the normal of the top surface, was kept relatively large. In this setup, relative motion between the observer and the object, i.e., rotation around a vertical axis, could be translated to depth variations in right angles on the top surface, making this solution effective. However, as the surface becomes more upright, this solution actually becomes less effective (see Fig. 1 of Wang et al. (2019) for an illustration).

Wang et al. (2019) conducted several experiments to explore the potential issue related to the right-angle solution of the bootstrap process. First, the original formulation of the right-angle solution assumed the identification and use of trackable texture elements to form right angles. However, the authors tested if the process still worked when they eliminated trackable points by using dynamic random dot stereograms (in which random texture was re-randomized in each frame and thus untrackable) while providing observers with trackable 3D structures that included inherent right angles. As predicted by the bootstrap process, performance improved

with large continuous perspective change. Results from this experiment showed that observers did not need trackable texture on the object to perform the bootstrap process, and instead, this process was performed in 3D relief space. Next, the task was tested after having eliminated the right angles inherent to the object using a symmetrical hexagon both with and without trackable texture elements. Participants still performed the task well with large rotation. This confirmed our suspicion that right angles were not the key to the bootstrap process.

Based on these findings, the authors presented an alternative formulation of the bootstrap process that used only two equidistant points on the object, originally being one leg of the right angle. Specifically, at the instance when the equidistant points are identified, the line formed between them, by definition, is on the image plane under scaled orthographical projection. Since observers are capable of perceiving distances on the image plane, the length of the line is known. With perspective change, such a line would no longer be on the image plane, and its depth dimension would be subject to relief scaling, making the 3D length of the line to be different from its actual length, as originally identified. Following the same logic as in the right-angle solution, one can extrapolate the correct scaling factor using the 2D projected length of this line through rotation.

Symmetry constrains the direction of slant

In addition to eliminating the inherent right angles in the objects, Wang et al. (2019) also explored the role of symmetry in slant perception as mediated by large continuous perspective change. They first used a series of asymmetric pentagonal objects and found that observers could not produce accurate slant judgments despite large continuous perspective change. Based on this finding, the authors inferred that this final failure might be because the direction of slant of an asymmetric object is ambiguous as it rotates. In a local frame of reference, the direction of slant is simply the direction at which there is a maximum increase in depth on the surface (e.g., Stevens 1983a, 1983b). In a geographical frame of reference (i.e., for geographical slant), this property can no longer be used to identify the direction of slant. Numerous studies have shown that observers readily detect the symmetry axes in a 2D shape oriented in a 3D environment (e.g., Bingham & Muchisky, 1993a, 1993b; Palmer, 1985; Pizlo et al., 2010; Sawada & Pizlo, 2008). In the experiments by Wang et al. (2019) using rectangular and hexagonal surfaces, a symmetry axis was aligned with the direction of slant. Therefore, in a subsequent experiment, they changed the direction of this symmetry axis in the hexagonal object by rolling the surface by 15°; judgments of slant accurately reflected the surface orientation along the symmetry axis rather than in the actual direction of slant. Thus, the bootstrap process worked, but the correctly

recovered scaling factor was then used to derive a slant estimate based on the symmetry axis.

We now present a detailed roadmap for implementation of the bootstrap process in the context of slant perception, starting with an SFM algorithm used to recover relief depth structure followed by the bootstrap to the correct scaling factor then used to produce slant estimation. Subsequently, we provide model simulation results using stimuli from Wang et al. (2018) and Wang et al. (2019) so we can compare our simulations with empirical results. This is the first study to formalize the full stratified process and present simulations of the collection of behavioral results that have been obtained. This is a test of the model itself, an evaluation of whether it is suitable for describing the achievement of accurate metric slant estimates using a large continuous perspective change. The model was developed based on the results of the initial experiments and is now tested by applying it to replicate results from a number of subsequent experiments.

Stratified recovery process

In this section, we illustrate the steps through which one could use the SFM algorithms and the bootstrap process to produce slant estimation. We first present an implementation of an existing structure-from-motion (SFM) algorithm, which was adopted from Lind (1996). This implementation takes two frames of 2D projected x-y coordinates of a 3D object moving in depth as input and produces a relief depth map of the object and other motion parameter estimates. In the second part, an implementation of the bootstrap process that utilizes the tracking of two equidistant points as initially revealed in Wang et al. (2019) was illustrated. Finally, we present a way through which the resulting scaling factor could be used to produce slant estimation. Throughout this paper, we use lower case letters, x , y , and z , to denote image coordinates or relief depth (as with z), and capital letters, X , Y , and Z , to denote physical coordinates.

Stage one: Recovery of relief depth map

There are many existing models that can recover relief depth based on two-frame motion (e.g., Shapiro, Zisserman, & Brady, 1995; Lind, 1996). Because SFM processes have been described and discussed at length in previous studies, we now briefly present a simplified version of the method proposed in Lind (1996). We note that the process is designed as a means for tracking invariant structure visible over perspective changes. Specifics of the formalization are for computational convenience, as was also made clear in Lind (1996). Again, as mentioned in the *Introduction*, the essence of the bootstrap process is tracking invariant 3D relief structure through SFM

transformations and using it to extrapolate 3D metric properties.

In the case of orthographic projection for SFM analyses, we decompose relative motion between an observer and an object in terms of a rotation of the image plane and a rotation around a unit axis that is in the image plane, as shown in Fig. 4 (Todd & Bressan, 1990; Shapiro et al., 1995). Lind (1996) denoted the rotational speed of the image plane as \dot{q} and the rotation around the axis in the plane as $\dot{\alpha}$. In addition, with scaled orthographic (or weak perspective) projection, there is a uniform shrinking or expansion of the 2D image, as a function of the speed in depth, \dot{z} . The unit axis that is in the image plane has a direction of δ , relative to the positive x-axis.

Given such decomposition in a scaled orthographic SFM analysis, one can produce estimates of the unknown parameters \dot{q} , \dot{z} , and δ with two 2D images based on the viewing of a moving and rigid 3D object. However, the $\dot{\alpha}$ parameter cannot be estimated due to the lack of degrees of freedom in the system. Instead, the depth within an unknown scaling factor ($\dot{\alpha}Z$) is estimated for each image location where Z denotes the distance to this 3D texture element and $\dot{\alpha}$ is the unknown scaling factor. Thus, $\dot{\alpha}$ is the unknown scaling factor in the bootstrap analysis and $\dot{\alpha}Z$ is the relief depth map produced by the SFM analysis.

By aligning the y-axis of a coordinate system in the image plane with the direction δ , the following equations can be obtained:

$$\dot{x} \approx -\dot{\alpha}Z + x\dot{z} - y\dot{q} \tag{1}$$

$$\dot{y} \approx y\dot{z} + x\dot{q} \tag{2}$$

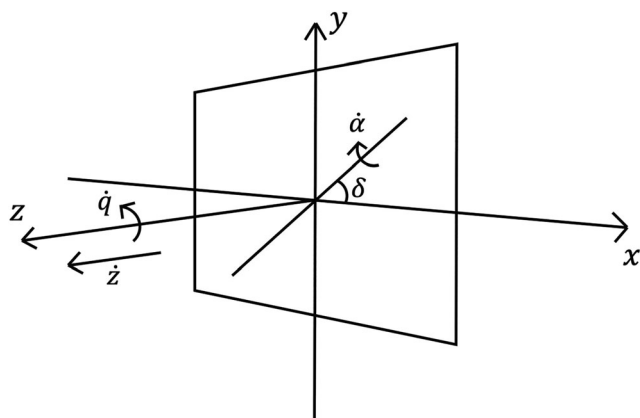


Fig. 4 A schematic demonstration of the motion decomposition in Lind (1996). \dot{q} is the angular velocity of the image plane, \dot{z} is the speed of translation in depth yielding a uniform shrinking or expansion of the image plane, δ is the direction of the unit axis in the image plane relative to the positive x-axis, and $\dot{\alpha}$ corresponds to the angular velocity of the rotation around the unit axis. The latter is the essential SFM transformation

where x and y denote the position in the image plane of a specific 3D texture element, and \dot{x} and \dot{y} are the velocity of this texture element across a two-frame apparent motion. Using \dot{z}_{est} and \dot{q}_{est} can be obtained through a linear process, which can then be combined with Eq. 1 to produce an $\dot{\alpha}$ scaled depth map:

$$\dot{\alpha}Z \approx -\dot{x} + x\dot{z}_{est} - y\dot{q}_{est} \tag{3}$$

For a more detailed discussion of this process, see Lind (1996).

Stage two: The bootstrap process

At this stage, we use the relief depth map, $\dot{\alpha}Z$, and information provided by the two-frame motion to produce an estimate of $\dot{\alpha}$ that would subsequently be used to produce an estimate of slant. The formulation proposed by Wang et al. (2019) starts with the identification of two points at t_0 , P_{1t_0} , $(x_{1t_0}, y_{1t_0}, z_{1t_0})$, and P_{2t_0} , $(x_{2t_0}, y_{2t_0}, z_{2t_0})$ that are equidistant to the observer. By definition, the length of the line formed between P_1 and P_2 is simply the distance between two points in the image plane (Fig. 5):

$$L = \sqrt{(x_{1t_0} - x_{2t_0})^2 + (y_{1t_0} - y_{2t_0})^2} \tag{4}$$

With relative motion between the observer and the object, for example, motion from rotation around an axis in the xy-plane, P_1 and P_2 will no longer be equidistant to the observer or be in the same frontoparallel plane. In addition, the 3D length of the line formed between these two points will be subject to relief scaling because this line is now lying in depth relative to the observer. Let $l(t)$ be the line's orthographically

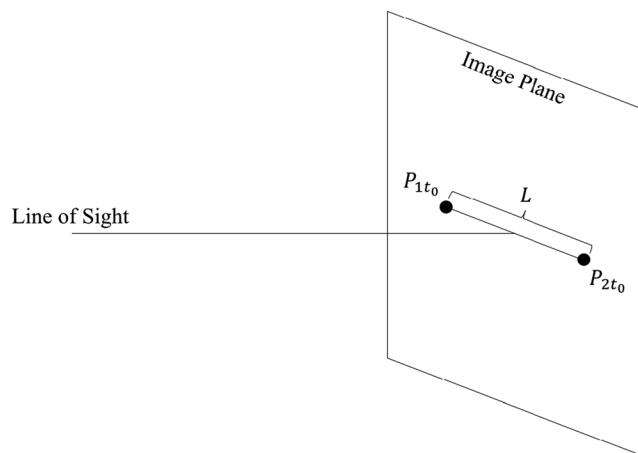


Fig. 5 The equidistant points at time t_0 . P_{1t_0} and P_{2t_0} are equidistant from the observer

projected 2D length in the image plane at any given time t :

$$l(t) = \sqrt{(x_{1t} - x_{2t})^2 + (y_{1t} - y_{2t})^2} \quad (5)$$

Assuming rigidity, we can establish a mapping between the 2D projected distance between the two points with the Euclidean 3D distance:

$$l = L \cos \alpha \quad (6)$$

Taking the derivative of Eq. 6, we have:

$$\dot{l} = L (-\sin \alpha) \dot{\alpha} \quad (7)$$

Combining Eq. 6 and Eq. 7, the angular velocity of rotation at any given time can be expressed as:

$$\dot{\alpha}(t) = \frac{\dot{l}}{\sqrt{L^2 - l(t)^2}} \quad (8)$$

One can subsequently substitute x and y coordinates of the two equidistant points to numerically derive an $\dot{\alpha}$ estimate. When noise is added to the system, measuring the projected length of the line could be variable. However, with an increasing amount of perspective change, the projected length would become increasingly smaller, and therefore is less affected by the addition of noise. Therefore, with a larger amount of rotation, $\dot{\alpha}$ estimations should also become more accurate.²

Because of the unknown scaling factor, it is difficult for the observer to discern whether there has been a sufficiently large continuous perspective change (just as we had discussed in the case of the right-angle solution). For the right-angle solution, Lind et al. (2014) argued that the observers would be informed when the line of sight bisects the right angle as symmetry and angle bisection are available in relief space. Naturally, bisection of a right angle corresponds to 45° of continuous perspective change. For the equidistant point solution, observers can evaluate the amount of rotation by comparing the change in length of the 3D relief distance between the two initially equidistant points. Norman, Todd, Perotti, and Tittle (1996) found that the threshold for comparing 3D lengths that had different orientation in space relative to the frontoparallel plane was between approximately 19% and 26%. A 20% difference in the 3D line length would correspond to 35.90° of rotation (i.e., change in tilt), whereas a 25% difference would correspond to 42.27° of rotation. This range of continuous perspective change is more flexible than the previous 45° criterion. In fact, depending on the experiments, veridical performance has

² The current implementation does not take into consideration motion in depth, i.e. *is presumed*. When there is motion-in-depth, the length variables can simply be scaled by that recovered from the SFM algorithm accordingly.

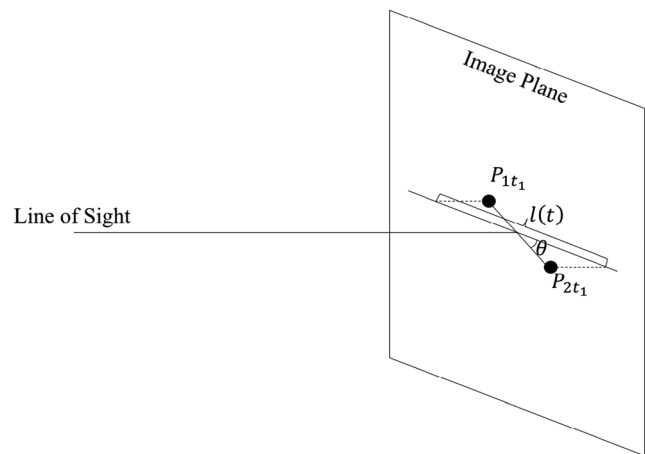


Fig. 6 The equidistant point setup at time t_1

been found empirically at either 35° or 45° of continuous perspective change (e.g., for the rectangle with cuboids stimuli in Experiment 2 of Wang et al. (2018) and Experiment 1 of Wang et al. (2019)).

Computationally, such change can be captured using the angle formed by the line between the two initially equidistant points at time t and the projection of that line in a frontoparallel plane. Because the length of the line should remain the same over rotation, using the projected 2D length of the line at time t and its original length we can calculate the cosine of the angle, θ , formed between the two lines (Fig. 6):

$$\cos \theta = \frac{l(t)}{L} \quad (9)$$

In this case, θ directly measures the amount of rotation that has occurred from the first frame. To provide a measure of quality that reflects the usefulness of rotation, we can use the following³:

$$quality = 4(\cos \theta - \cos^2 \theta) \quad (10)$$

This measure of quality was computed so that it monotonically increased for θ ranging between 0° and 60°. Based on the criterion angles established earlier, we can compute the corresponding quality measure to evaluate the amount of rotation that has occurred.

³ Perceptually, evaluating whether there has been a sufficiently large amount of perspective change relies on the change in length of the line formed between two equidistant points. Computationally, evaluating whether sufficiently large rotation has occurred relies on the value of θ . We formulated and used this particular measure of quality simply because it increases monotonically rather than oscillating around 1 as does, for instance, $\cos \theta$. Other measures of quality that are based on θ would be equally suitable.

Stage three: Slant estimation

Given the relief depth map from Stage 1 and the recovered scaling factor from Stage 2, the final step is to recover the perceived slant. Based on discussions in the previous section, we adopted a geographical frame of reference, where slant is defined as the angle between the surface normal and the normal of the ground surface, i.e. unit vector $[0, 1, 0]$, and tilt is defined as angle between the projection of the surface normal on the ground plane (xz -plane) and the direction of line of sight, i.e. unit vector $[0, 0, 1]$ (Fig. 3). For a planar relief surface centered at the origin, its surface equation can be expressed as the following:

$$z = \hat{\alpha}Z = ax + by \tag{11}$$

where z is the relief depth, which is equivalent to its actual depth, Z , scaled by the scaling factor $\hat{\alpha}$. Since we have already obtained z and $\hat{\alpha}$ in the first two stages, we can again use linear processes to acquire coefficients a and b . Subsequently, the plane equation can be expressed as:

$$ax + by - \hat{\alpha}Z = 0 \tag{12}$$

The surface normal is therefore $(a, b, -\hat{\alpha})$. We can then find the angle between the surface normal and unit vector $[0, 1, 0]$ using cross and dot products:

$$\sin(\sigma) = \frac{\sqrt{\hat{\alpha}^2 + a^2}}{\sqrt{a^2 + b^2 + \hat{\alpha}^2}} \tag{13}$$

$$\cos(\sigma) = \frac{b}{\sqrt{a^2 + b^2 + \hat{\alpha}^2}} \tag{14}$$

Dividing Eq. 13 by Eq. 14 yields a simpler expression of σ :

$$\tan(\sigma) = \frac{\sqrt{\hat{\alpha}^2 + a^2}}{b} \tag{15}$$

Because $\hat{\alpha}$ is known from the bootstrap model, slant can be estimated.

Additionally, Wang et al. (2019) found that when the direction of slant is not aligned with the direction of the symmetry axis of an object, perception of slant tended to reflect that along the symmetry axis. To address the issue of slant direction, we will use a second method to derive slant estimates. First, for symmetrical objects, we assume that the observers can correctly identify the symmetry axis of an object. There are many studies that model the perception of symmetry and skewed symmetry axes (see, e.g., Li, Sawada, Shi, Kwon, & Pizlo, 2011). With the symmetry axis on the object, along with the object’s normal, we can find a plane that passes through both vectors. We computed slant to be the angle between the normal of that surface and the unit vector $[0, 1, 0]$. For asymmetrical objects, namely the pentagon surface with

tetrahedrons used in Experiment 3 of Wang et al. (2019), a random point would be chosen along the edge of the surface in our simulation and be connected to the pentagon’s top vertex as the direction of slant to reflect the uncertainty of the direction of slant. Therefore, this could be a test of whether in fact observers exhibited random slant directions in their judgments when the surface lacks symmetry.

Model simulation

We now look at how model predictions compared to actual human performance. We performed simulations of four experiments with three different types of objects that we have used in slant perception studies (Wang, Lind, & Bingham, 2018, 2019). The first type of object was a rectangular surface with nine cuboids on top in a grid, as used in Experiment 2 of Wang et al. (2018). This is to test if the proposed stratified process could predict human performance, i.e., whether sufficiently large continuous perspective change would allow accurate slant estimation. This object contained the most structure (right angles and symmetry) compared to the others to be used, so it provided a baseline test of the effectiveness of the algorithms. The second type of object was hexagonal with tetrahedrons on top at random locations, as in Experiment 2 of Wang et al. (2019). Additionally, we used this object with changes in the orientation of the symmetry axis, which was then used to derive alternative slant estimates, as in Experiment 4 of Wang et al. (2019). These two simulations tested the claim in Wang et al. (2019) that observers used the direction of the symmetry axis to determine the slant direction. Finally, we used an asymmetric pentagonal surface with randomly placed tetrahedrons to simulate the results in Experiment 3 of the same study. In this simulation, slant was computed using a random reference line on the object as the direction of slant. This simulation tested the claim in Wang et al. (2019) that with a lack of symmetry axis, observers randomly chose the direction of slant.

We did not simulate performance with strictly planar surfaces because the lack of non-coplanar points would prevent the SFM process from recovering the unknown motion parameters in the beginning of the stratified process. Specifically, the planar surface would produce one less linearly independent equation than the number of unknown motion parameters for the system to solve. Recall from Stage 1 of the stratified process that this means that either z_{est} or q_{est} was still remain unknown after the SFM process. However, the current experimental setup lacks z and q motion components, allowing the recovered relief depth map to be usable. This does not apply to the human visual system and the resulting depth map cannot be confidently used for the subsequent bootstrap process. To explore the effects of the lack of non-

coplanar points on the bootstrap process, we incorporated motion-in-depth, i.e., z , in a simulation, comparing the effectiveness of the bootstrap process in recovering the unknown scaling factor with a rectangular planar surface and the same surface with additional cuboids. We simulated a continuous rotation of 90° , combined with a translation along the z -axis of 4 cm during the rotation. We identified 50 pairs of equidistant points and computed the estimated scaling factor for each pair across rotation. The final result at each tilt was the mean of those pairs. Figure 7 shows the simulation results. The bootstrap process could successfully recover the unknown scaling factor for the surface with additional 3D structure in the presence of motion-in-depth, but not for the strictly planar surface.

Experimental stimuli and setup

The three different types of objects used both in human experiments and in simulations are illustrated in Fig. 8, including the rectangular surface with nine cuboids located in a 3-by-3 grid, the hexagonal surface with nine tetrahedrons at random locations, and the asymmetrical pentagonal surface with nine tetrahedrons at random locations, together with the manipulation of the orientation of the symmetry axis of the hexagonal surface. In addition, as shown, we rotated the hexagonal surface around its surface normal by 15° to manipulate the orientation of its symmetry axis. All surfaces had the same width, 10 cm, and heights, 8 cm, 10 cm, or 12 cm. The cuboids and tetrahedrons had heights of 0.55 cm and base lengths of 1 cm.

To test the effects of large continuous perspective change, there were five different rotation amounts, from 25° to 65° with a 10° increment. Objects were presented using half rotation. For instance, with a 25° rotation, the object, started facing the observer, at a tilt of 0° , first rotated through 12.5° to one side, rotated back to the starting orientation, and then rotated through 12.5° to the other side. Within each rotation

amount, there was a total of 24 different slant angles, from 27° to 73° with a 2° increment. Finally, for three consecutive slant angles, the height of the surface pseudo-randomly varied among 8 cm, 10 cm, and 12 cm. Therefore, for each shape condition, there was a total of 120 trials (five different rotation amounts times 24 different slant angles). To match sample size of different experiments, we simulated 12 participants' data using the rectangular strictly planar surface experiment, 10 using the rectangular surface with cuboids experiment, 11 using the hexagonal surface with tetrahedrons experiment, and 10 using the asymmetric pentagonal surface with tetrahedrons experiment. Stimuli were presented using the monocular SFM information in the actual experiment. The objects consisted of random dots, located 9 cm behind and screen. They were back projected onto the screen from a projection point 76.2 cm in front of the screen, yielding a viewing distance of 85.2 cm.

Simulation methods

Each simulated object consisted of the same random texture points as used in the corresponding human experiment, with a texture density of 26.67 points per cm^2 . We first constructed the objects using these points, rotated the objects, and projected points on the object to the screen, generating x and y coordinates for each point. Coordinates were then scaled by the viewing distance, which was 85.5 cm. Gaussian noise with a standard deviation of 0.002 was introduced to the coordinates to reflect random error associated with motion measurement. Each frame had different noise seeds.

We generated inputs to the model for each trial. Within each trial, a set of 50 samples each consisting of two equidistant points found on the surface were identified. The model then tracked these samples across the entire range of rotation, producing a quality measure, $\hat{\alpha}$, and the surface plane

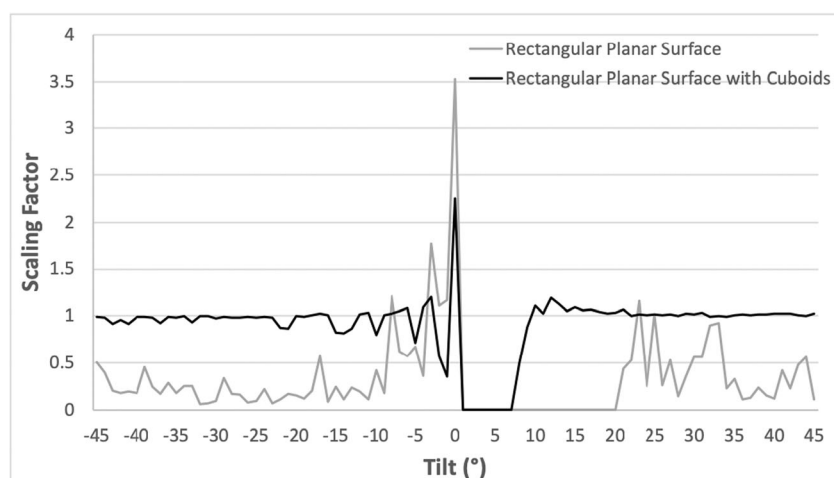


Fig. 7 Recovered scaling factors through the bootstrap process for a strictly rectangular planar surface and the same surface with nine cuboids plotted as a function of change in tilt. The correct scaling factor is 1

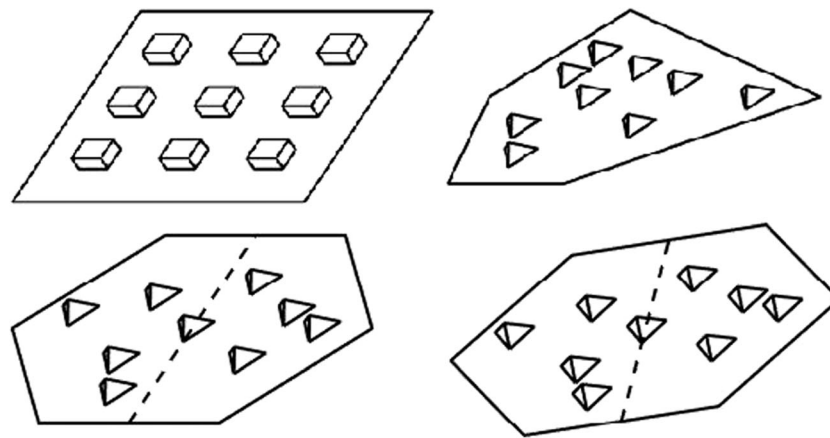


Fig. 8 Schematic illustration of surfaces used in experimental displays and model simulations: rectangular surface with nine cuboids in a three-by-three grid (top left), asymmetrical pentagonal surface with nine randomly located tetrahedrons (top right), hexagonal surface with randomly located tetrahedrons (bottom left), and the same hexagonal surface with

the symmetry axis rotated around the surface normal by 15° (bottom right). For the hexagonal objects, dashed lines represent the direction of symmetry axis. All objects presented have a slant of 45° and a tilt of -32.5°

coefficients. The model first checked among the samples whether the respective quality measures satisfied a criterion. The quality criterion was set based on a 19% 3D distance discrimination threshold (Norman et al., 1996). We first identify samples that had a quality measure that exceeded the criterion. If such samples exist, we used the median $\hat{\alpha}$ estimates at the tilt where the quality measure was above criterion for those samples, and the final $\hat{\alpha}$ estimate was the mean of the medians. Alternatively, if no sample satisfied the quality criterion, we simply used the mean of the medians of all samples across all tilts as the final $\hat{\alpha}$ estimation.

Furthermore, as numerous studies have suggested, human observers tended to perceive more upright slants to be closer to the frontoparallel plane (and thus, as more slanted) (Todd, Thaler, & Dijkstra, 2005; Norman et al., 2009; Durgin, Li, & Hajnal, 2010; Saunders & Chen, 2015). In addition, Cherry and Bingham (2018) reported a similar bias towards shallower slants, where observers tended to perceive smaller slants to be closer to the horizontal ground surface. According to the authors, such a bilateral bias could be attributed to the presence of vertical (e.g., walls) and horizontal (e.g., table) references in the experiment. Similarly, Saunders and Chen (2015) argued that for computer-generated displays, presence of the vertical computer screen could generate a frontal bias. To represent these biases, we introduced additional noise to the final $\hat{\alpha}$ estimate in the model that was proportional to the derived slant when no sample had a quality measure that satisfied the criterion. Such noise would make large slants more frontoparallel and small slants more horizontal.

Finally, based on findings of Wang et al. (2019), the direction of slant could also vary depending on the presence or absence of mirror symmetry in the object. To capture this, for the pentagonal objects, we used the symmetry axis as its slant direction. For the pentagonal objects, we picked a

random point along the bottom edge of the object and connected it with its top vertex and used this as its slant direction. We varied the slant direction for pentagonal objects for every trial.

Data analysis

To analyze and compare model prediction and actual human performance, we used the same analysis protocol as in Wang, Lind, and Bingham (2018, 2019). Specifically, we used linear regression, regressing predicted slant onto actual slant, and regression slopes and intercepts for each rotation amount and each participant as measures of performance. A veridical judgment of slant within a rotation condition entails a regression slope of 1 and a regression intercept of 0. To compare model prediction with human performance, we used mixed-design analysis of variance (ANOVA) on regression slopes, intercepts, and r^2 with one within-subject factor of rotation amount (five levels) and one between-subject factor of type of results (two levels, human performance or model prediction).

Simulation results

For the rectangular surface with cuboids, regression slopes and intercepts for human performance and model simulation are shown in Fig. 9. Mauchly's test of sphericity showed that the sphericity assumption was violated ($\chi^2(9) = 27.55$, $p = 0.001$). With Greenhouse-Geisser correction, ANOVA on regression slopes showed that there was a significant main effect of rotation amount ($F(2.50, 55.00) = 7.00$, $p = 0.001$, $\eta_p^2 = 0.24$). There was no significant effect of the type of results ($p > 0.9$, $\eta_p^2 = 0.00$) or a significant interaction effect between the two ($p > 0.7$, $\eta_p^2 = 0.015$). For regression intercepts, the

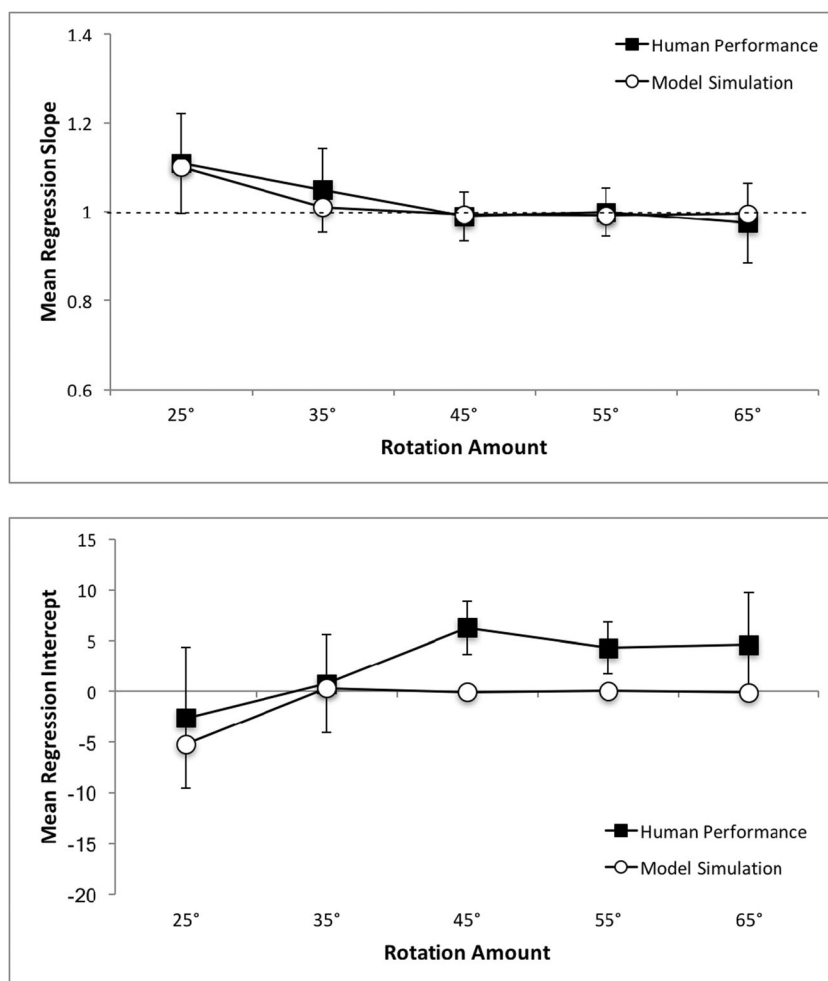


Fig. 9 Mean regression slopes (top) and intercepts (bottom) for human performance and model simulation using the rectangular surface with cuboids. Error bars represent 95% confidence intervals around the mean,

calculated for repeated-measures designs (Cousineau, 2005; with correction by Morey, 2008)

sphericity assumption was also violated ($\chi^2(9) = 35.65, p < 0.001$). ANOVA on regression intercepts showed that there was a significant main effect of rotation ($F(2.40, 52.72) = 7.44, p = 0.001, \eta_p^2 = 0.25$) after Greenhouse-Geisser correction. There was neither a significant main effect of source of data ($p > 0.2, \eta_p^2 = 0.05$) nor a significant interaction between the two factors ($p > 0.1, \eta_p^2 = 0.07$). As Fig. 9 shows, human performance and model prediction are identical for regression slopes. Although regression intercepts were a little higher for human performance than for model prediction, this difference was not statistically significant.

Based on comparisons of human and simulation results for the cuboids display, one can see that the equidistant points bootstrap process works well. In particular, slant estimation was inaccurate at 25° of rotation. With 35° rotation, model performance improved a little but did not reach the veridical level. Finally, at 45° and beyond, slant estimations became accurate. Results from this simulation had two implications.

First, the stratified process of deriving slant estimates using the equidistant points implementation is suitable for describing performance in 3D slant perception. As the results suggested, model performance did improve with an increasing amount of rotation and remained at a steady level once rotation reached and went beyond 45°. Secondly, the additional noise introduced to the final estimates of the scaling factor when the quality measure did not exceed threshold accurately reflected the tendency of human perception of 3D slant, namely that people tended to overestimate large slant (more upright slants were perceived to be even more upright).

Next, using the same method, we generated model simulations of the hexagonal surface with tetrahedrons at random locations. Figure 10 shows the mean regression slopes and intercepts for human and model performance. Again, we saw identical performance in model and data. For regression slopes, the sphericity assumption was violated ($\chi^2(9) = 30.68, p < 0.001$). With Greenhouse-Geisser correction, ANOVA showed that there was a significant main effect of rotation

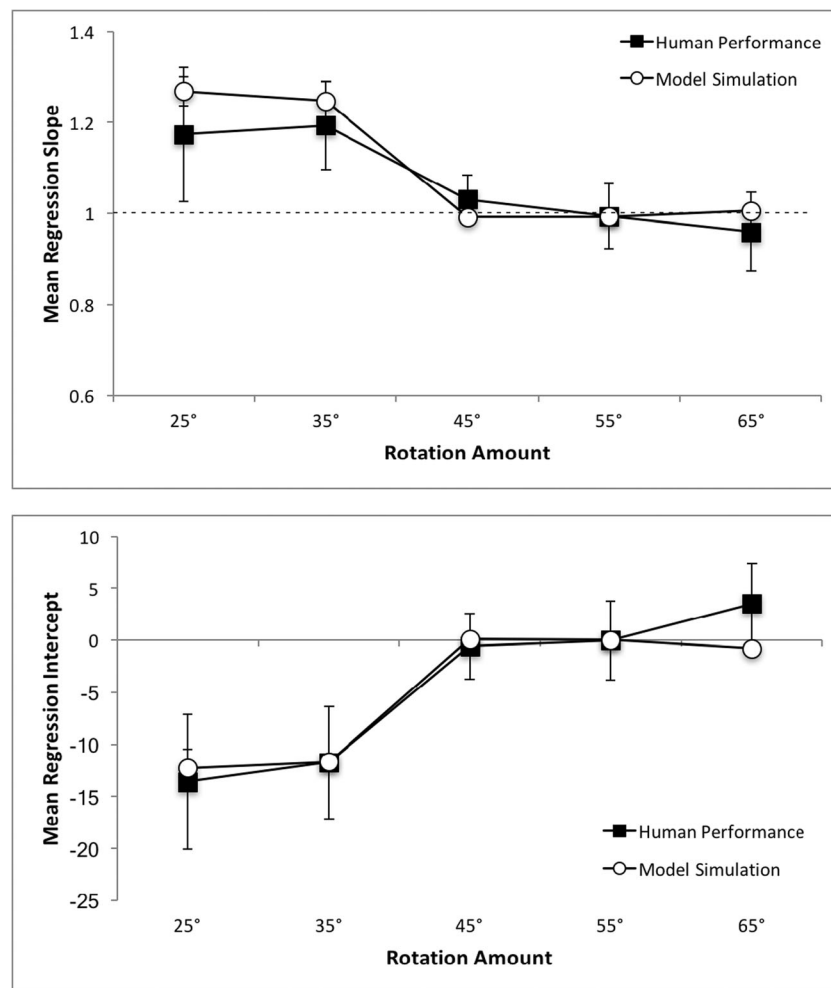


Fig. 10 Mean regression slopes (top) and intercepts (bottom) for human performance and model simulation using the hexagonal surface with random tetrahedrons. Error bars represent 95% confidence intervals around

amount ($F(2.35,47.04) = 31.59, p < 0.001, \eta_p^2 = 0.61$). However, there was neither a main effect of the type of results ($p > 0.5, \eta_p^2 = 0.02$) nor a significant interaction effect between the two factors ($p > 0.2, \eta_p^2 = 0.064$). For regression intercepts, the sphericity assumption was violated ($\chi^2(9) = 19.73, p < 0.05$). With Greenhouse-Geisser correction, ANOVA showed that there was a significant main effect of rotation amount ($F(2.79,55.80) = 42.01, p < 0.001, \eta_p^2 = 0.68$). There was neither a main effect of type of results ($p > 0.8, \eta_p^2 = 0.002$) nor a significant interaction effect between the two factors ($p > 0.3, \eta_p^2 = 0.05$). As can be seen from Fig. 10, the model predicted the pattern of performance for the hexagonal objects very well. The model successfully predicted veridical performance at 45° of rotation and beyond.

Next, we performed simulations for hexagonal displays with a 15° roll. In our simulation, to explore the role of symmetry, we computed slants based on the direction of the symmetry axis instead of using the surface normal. Figure 11 shows the mean

the mean, calculated for repeated-measures designs (Cousineau, 2005; with correction by Morey, 2008)

regression slopes and intercepts for human performance and model simulation. Mauchly's test of sphericity showed a violation of the sphericity assumption for regression slopes ($\chi^2(9) = 21.87, p = 0.01$). With Greenhouse-Geisser correction, there was a significant main effect of rotation amount ($F(2.63,47.39) = 25.03, p < 0.001, \eta_p^2 = 0.62$). There was no significant main effect of the source of data ($p > 0.7, \eta_p^2 = 0.007$) or a significant interaction effect between the two factors ($p > 0.5, \eta_p^2 = 0.041$). For regression intercepts, there was also a significant main effect of rotation amount ($F(4,72) = 28.88, p < 0.001, \eta_p^2 = 0.62$). Again, there was neither a not a significant main effect of source of data ($p > 0.8, \eta_p^2 = 0.005$) nor a significant interaction effect ($p > 0.2, \eta_p^2 = 0.076$).

As shown in Fig. 11, model simulation produced the same pattern as human performance both in terms of regression slopes and intercepts. Wang et al. (2019) suggested that observers judged slant using the symmetry axis to determine the direction of slant. They subsequently regressed slant

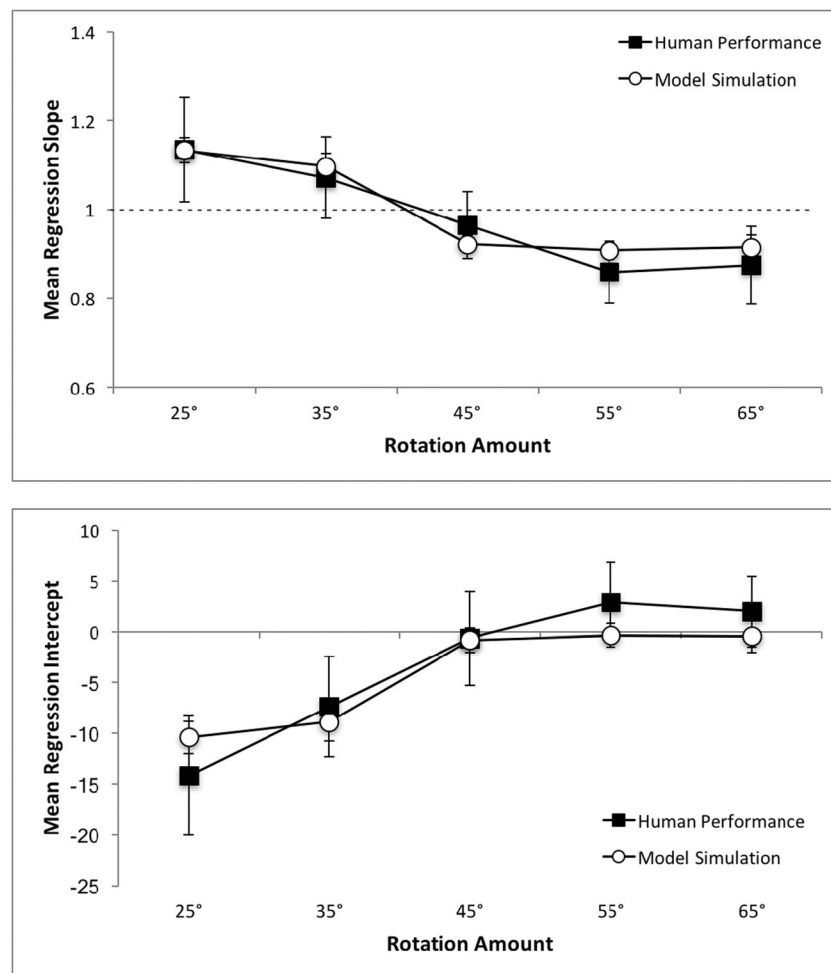


Fig. 11 Mean regression slopes (top) and intercepts (bottom) for human performance and model simulation using the hexagonal surface with random tetrahedrons with a 15° rotation of the surface around its normal.

Error bars represent 95% confidence intervals around the mean, calculated for repeated-measures designs (Cousineau, 2005; with correction by Morey, 2008)

judgments onto slants that were computed along the direction of the object's symmetry axis. With this, the dependent measures reached veridical level (i.e., slope of 1 and intercept of 0). In our simulation, after using the recovered relief scaling factor to adjust recovered depth values, we computed slant along the direction of the symmetry axis and regressed slant estimates from the model onto actual (correct) slants. This produced results that are equivalent to human performance when it is compared to actual (correct) slants. This provided additional support for the claim that human observers did in fact used the object's symmetry axis as the direction of slant and that they could still recover the correct scaling factor even though symmetry was perturbed.

Finally, we simulated results from the pentagonal surface with tetrahedrons. Wang et al. (2019) suggested that poor performance in this case was due to the lack of specification of the direction of slant. To simulate such an effect, for each trial, we randomly chose a point along the bottom edge of the pentagonal surface and formed an axis connecting that point to the top vertex of the surface. This yielded poorly constrained and

variable estimates of the slant direction. Figure 12 shows the mean regression slopes and intercepts. Again, sphericity assumption was violated for regression slopes ($\chi^2(9) = 19.33$, $p < 0.05$). With Greenhouse-Geisser correction, ANOVA showed that there was a significant main effect of rotation ($F(2.68, 48.20) = 10.65$, $p < 0.001$, $\eta_p^2 = 0.37$). However, there was neither a significant main effect of source of data ($p > 0.4$, $\eta_p^2 = 0.033$) nor a significant interaction effect between the two tested factors ($p > 0.05$, $\eta_p^2 = 0.12$). For regression intercepts, the sphericity assumption was violated ($\chi^2(9) = 21.69$, $p < 0.05$). With Greenhouse-Geisser correction, there was a significant main effect of rotation ($F(2.60, 46.81) = 8.77$, $p < 0.001$, $\eta_p^2 = 0.33$). There was no effect of source of data ($p > 0.1$, $\eta_p^2 = 0.08$) but there was a significant interaction effect between the two factors ($F(2.60, 46.81) = 3.50$, $p < 0.05$, $\eta_p^2 = 0.16$).

As seen from Fig. 12, regression slopes from model simulations again replicated those from human performance. There was a trend toward veridical performance over increasing amounts of rotation despite the random estimates of the

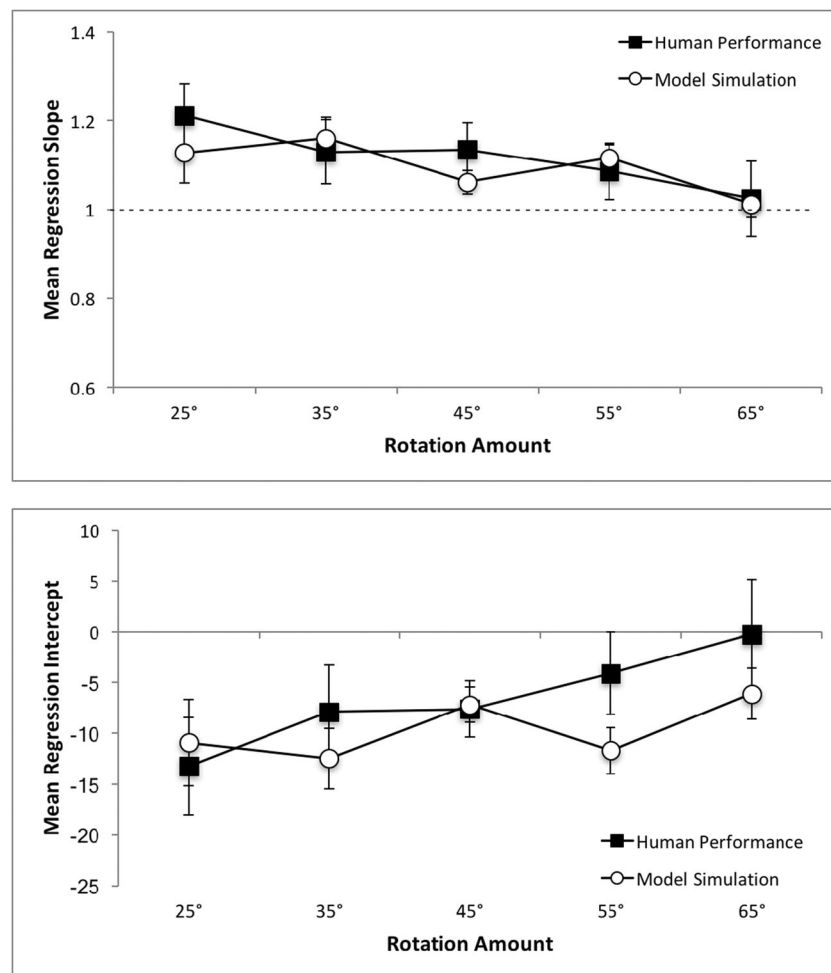


Fig. 12 Mean regression slopes (top) and intercepts (bottom) for human performance and model simulation using the pentagonal surface with random tetrahedrons. Error bars represent 95% confidence intervals

direction of slant. This was true of both regression slopes and intercepts. However, there was a discrepancy between model simulation and human performance in intercepts at 55° of rotation, which contributed to a significant interaction effect. As seen from Fig. 12, the model yielded smaller intercepts than human results. Overall, the significance of this simulation is that it provided evidence for arguments made in Wang et al. (2019) that for the pentagonal surfaces, due to the lack of symmetry, observers chose the direction of slant randomly for each trial. Because the model was able to produce accurate estimates of the unknown scaling factor, this suggested that observers in the pentagonal experiment were able to recover the correct scaling factor as well.

General discussion

We presented a thorough illustration of how the bootstrap process was implemented in addition to providing an alternative solution to the process itself that was more suitable for 3D

around the mean, calculated for repeated-measures designs (Cousineau, 2005; with correction by Morey, 2008)

slant perception. We simulated experiments from Wang, Lind, and Bingham (2018, 2019), including rectangular surfaces with cuboids, hexagonal surfaces with randomly distributed tetrahedrons, both with symmetry axis aligned with and not aligned with the direction of slant, as well as pentagon surfaces with tetrahedrons. Based on simulation results, we can see that the model has an extremely high predictive power, resulting in identical regression slopes and intercepts for various rotations across different slants.

Due to the high predictive power of the model, several crucial findings could be confirmed. First and foremost, the bootstrap process does not require right angles. The original solution presented in Lind et al. (2014) utilized the tracking of right angles across rotation. However, the essence of the bootstrap process is to track invariance over transformation. The right angles identified in the previous solution are equivalent in relief and physical space at the time when they are identified, enabling one to track its transformation across perspective change as a result of the unknown scaling in the relief space and to extrapolate such scaling factor. The alternative

solution as presented in the current study uses portion of the right angle, namely the line formed between two equidistant points, and track it instead. This solution is simpler and more elegant, avoiding the challenges that arise when one started to introduce slant to the picture. As simulations showed, this method is equally powerful of obtaining the correct scaling factor with large continuous perspective change. Interestingly, the fundamental ideas behind the bootstrap process (invariance over transformation) is congruent with the idea of *congruence under motion* developed by Lappin and colleagues (Lappin & Fuqua, 1983; Lappin & Love, 1992; Lappin & Ahlström, 1994).

Secondly, results from the hexagonal objects with a 15° roll as well as the pentagonal objects validated the claim from Wang et al. (2019) that symmetry of the objects aided the recovery of slant, not necessarily the recovery of the correct scaling factor. For the hexagonal objects, slant estimations were derived based on the direction of the symmetry axis and it replicated human performance. This indicated that observers did indeed utilize the direction specified by hexagons' symmetry axis to produce slant judgment. Similarly, slant estimations for the pentagonal objects were based on a random slant direction and it also replicated human performance. Because the bootstrap process could recover the correct scaling factor at 45° of rotation, poor model performance was primarily due to the randomness in the direction of slant. Human performance in this experiment did exhibit the same pattern of results, suggesting that human observers were likely to utilize the same strategy.

Third, the *post hoc* noise added to the scaling factor as a function of the initially derived slant when the quality measure failed to pass usable threshold helped the simulation results to exactly replicated human results confirmed that there was indeed a binary bias towards slant perception. According to Durgin, Li, and Hajnal (2010), there was a tendency for human observers to view slants to be more frontoparallel. Such a noise added to the scaling factor reflected such a tendency when the bootstrap's quality measure did not reach to a satisfiable level. This, to some extent, confirmed the bias as suggested by Durgin et al. In addition, based on both human performance and model simulation, we can see that such a bias can indeed be eliminated when there was a sufficiently large continuous perspective change.

Finally, the effectiveness of the entire stratified process of recovering slant also suggested that slant perception, at least 3D slant perception, may not be the basis for 3D shape perception. The modeling effort itself started from the recovery of relief structures based on four non-coplanar points across a two-frame apparent motion. This process itself was used to recover 3D shapes, not slant. In fact, the SFM algorithm cannot even operate with only coplanar points, which are what slants in a traditional sense is all about. If perceiving slant is the basis for perceiving 3D shapes, then the process should

start with somehow recovering the slant – meaning that the process should work with even the simplest planar surfaces. This is not the case. An overall shape of the object has to be recovered before one is able to recover slant. This suggests, but not completely substantiates, the fact that slant may just be a certain property of a 3D object, which is consistent with findings from other studies (i.e., Lappin, Norman, & Phillips, 2011). The surface of an object could have a certain orientation relative to the ground surface. Such orientation could be specified through the symmetry axis if the surface is symmetrical. Alternatively, if the orientation of the surface was underspecified, then observers would just randomly pick a direction that works out for them.

Open Practice Statement Data and materials for all experiments are available upon request to the corresponding author, and none of the experiments was preregistered.

References

- Beck, J., & Gibson, J.J. (1955). The relation of apparent shape to apparent slant in the perception of objects. *Journal of Experimental Psychology* 50(2), 125-133.
- Bingham, G. P., & Muchisky, M. M. (1993a). Center of mass perception and inertial frames of reference. *Perception & Psychophysics*, 54(5), 617-632.
- Bingham, G. P., & Muchisky, M. M. (1993b). Center of mass perception: Perturbation of symmetry. *Perception & Psychophysics*, 54(5), 633-639.
- Cherry, O. C., & Bingham, G. P. (2018). Searching for invariance: Geographical and optical slant. *Vision Research*, 149, 30-39.
- Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, 1(1), 42-45.
- Domini, F., & Caudek, C. (2013). Perception and action without veridical metric reconstruction: an affine approach. In S. Dickinson, & Z. Pizlo (Eds.) *Shape Perception in Human and Computer Vision* (pp. 285-298). Springer London.
- Durgin, F. H., Li, Z., & Hajnal, A. (2010). Slant perception in near space is categorically biased: Evidence for a vertical tendency. *Attention, Perception, & Psychophysics*, 72(7), 1875-1889.
- Gibson, J. J. (1950). The perception of visual surfaces. *The American Journal of Psychology*, 63(3), 367-384.
- Gibson, J.J., & Comsweet, J. (1952). The perceived slant of visual surfaces – optical and geographical. *Journal of Experimental Psychology*, 44(1), 11-15.
- Hoffman, D.D., & Richards, W.A. (1984). Parts of recognition. *Cognition*, 18, 65-96.
- Kaiser, P.K. (1967). Perceived shape and its dependency on perceived slant. *Journal of Experimental Psychology*, 75(3), 325-353.
- Koenderink, J. J., & Van Doorn, A. J. (1991). Affine structure from motion. *JOSA A*, 8(2), 377-385.
- Koffka, K. (1935). *Principles of Gestalt Psychology*. New York, NY: Harcourt Brace.
- Lappin, J. S., & Fuqua, M. A. (1983). Accurate visual measurement of three-dimensional moving patterns. *Science*, 221(4609), 480-482.

- Lappin, J. S., & Love, S. R. (1992). Planar motion permits perception of metric structure in stereopsis. *Perception & Psychophysics*, *51*(1), 86–102.
- Lappin, J. S., & Ahlström, U. B. (1994). On the scaling of visual space from motion—in response to Pizlo and Salacfa-Golyska. *Perception & psychophysics*, *55*(2), 235–242.
- Lappin, J. S., Norman, J. F., & Phillips, F. (2011). Fechner, information, and shape perception. *Attention, Perception, & Psychophysics*, *73*(8), 2353–2378.
- Lee, Y. & Bingham, G.P. (2010). Large perspective changes ($>45^\circ$) yield perception of metric shape that allows accurate feedforward reaches-to-grasp and it persists after the optic flow has stopped! *Experimental Brain Research*, *204*, 559–573.
- Lee, Y. L., Crabtree, C. E., Norman, J. F., & Bingham, G. P. (2008). Poor shape perception is the reason reaches-to-grasp are visually guided online. *Perception & psychophysics*, *70*(6), 1032–1046.
- Lee, Y.L., Lind, M., Bingham, N. & Bingham, G.P. (2012). Object recognition using metric shape. *Vision Research*, *69*, 23–31.
- Li, Y., Sawada, T., Shi, Y., Kwon, T., & Pizlo, Z. (2011). A Bayesian model of binocular perception of 3D mirror symmetrical polyhedra. *Journal of Vision*, *11*(4), 11, 1–20.
- Lind, M. (1996). Perceiving motion and rigid structure from optic flow: A combined weak-perspective and polar-perspective approach. *Perception & Psychophysics*, *58*(7), 1085–1102.
- Lind, M., Lee, Y. L., Mazanowski, J., Kountouriotis, G. K., & Bingham, G. P. (2014). Affine operations plus symmetry yield perception of metric shape with large perspective changes ($\geq 45^\circ$): Data and model. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 83.
- Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*(2), 61–64.
- Norman, J. F., Todd, J. T., Perotti, V. J., & Tittle, J. S. (1996). The visual perception of three-dimensional length. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(1), 173–186.
- Norman, J. F., Todd, J. T., Norman, H. F., Clayton, A. M., & McBride, T. R. (2006). Visual discrimination of local surface structure: Slant, tilt, and curvedness. *Vision Research*, *46*(6–7), 1057–1069.
- Norman, J. F., Crabtree, C. E., Bartholomew, A. N., & Ferrell, E. L. (2009). Aging and the perception of slant from optical texture, motion parallax, and binocular disparity. *Perception & Psychophysics*, *71*(1), 116–130.
- Palmer, S. E. (1985). The role of symmetry in shape perception. *Acta Psychologica*, *59*(1), 67–90.
- Pizlo, Z., Sawada, T., Li, Y., Kropatsch, W. G., & Steinman, R. M. (2010). New approach to the perception of 3D shape based on veridicality, complexity, symmetry and volume. *Vision Research*, *50*(1), 1–11.
- Sakata, H., Tsutsui, K.-I., & Taira, M. (2005). Toward an understanding of the neural processing for 3D shape perception. *Neuropsychologia*, *43*, 151–161.
- Sawada, T., & Pizlo, Z. (2008). Detection of skewed symmetry. *Journal of Vision*, *8*(5): 14, 1–18.
- Saunders, J. A., & Chen, Z. (2015). Perceptual biases and cue weighting in perception of 3D slant from texture and stereo information. *Journal of Vision*, *15*(2): 14, 1–24.
- Saunders, J. A., & Knill, D. C. (2001). Perception of 3D surface orientation from skew symmetry. *Vision Research*, *41*(24), 3163–3183.
- Shapiro, L.S., Zisserman, A., & Brady, M. (1995). 3D motion recovery via affine epipolar geometry. *International Journal of Computer Vision*, *16*(2), 147–182.
- Stevens, K.A. (1983a). Slant-tilt: The visual encoding of surface orientation. *Biological Cybernetics*, *46*, 183–195.
- Stevens, K. A. (1983b). Surface tilt (the direction of slant): A neglected psychophysical variable. *Perception & Psychophysics*, *33*(3), 241–250.
- Todd, J.T. (2004). The visual perception of 3D shape. *Trends in Cognitive Sciences*, *8*(3), 115–121.
- Todd, J.T., & Bressan, P. (1990). The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Perception and Psychophysics*, *48*(5), 419–430.
- Todd, J. T., & Norman, J. F. (1991). The visual perception of smoothly curved surfaces from minimal apparent motion sequences. *Perception & Psychophysics*, *50*(6), 509–523.
- Todd, J. T., Oomes, A. H., Koenderink, J. J., & Kappers, A. M. (2001). On the affine structure of perceptual space. *Psychological Science*, *12*(3), 191–196.
- Todd, J. T., Thaler, L., & Dijkstra, T. M. (2005). The effects of field of view on the perception of 3D slant from texture. *Vision Research*, *45*(12), 1501–1517.
- Wagner, M. (1985). The metric of visual space. *Perception & Psychophysics*, *38*(6), 483–495.
- Wallach, H., & Moore, M. E. (1962). The role of slant in the perception of shape. *American Journal of Psychology*, *75*, 285–293.
- Wang, X. M., Lind, M., & Bingham, G. P. (2018). Large Continuous Perspective Change with Noncoplanar Points Enables Accurate Slant Perception. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(10), 1508–1522.
- Wang, X.M., Lind, M., & Bingham, G.P. (2019). Symmetry mediates the bootstrapping of 3D relief slant to metric slant. *Attention, Perception, & Psychophysics*. <https://doi.org/10.3758/s13414-019-01859-5>.
- Welchman, A.E., Deubelius, A., Conrad, V., Bültöff, H.H. & Kourtzi, Z. (2005). 3D shape perception from combined depth cues in human visual cortex. *Nature Neuroscience*, *8*, 820–827.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.