Research Note

On the Relationship Between Deformation and Perceived Surface Slant

TOM C. A. FREEMAN,*t MIKE G. HARRIS,§ TIM S. MEESE

Received 27 December 1993; in revised form 10 June 1994; in final form 15 March 1995

A compelling impression of surface slant is produced by random dot displays depicting deformation and translation alone. A simple model of slant estimation based upon deformation is shown to capture quantitatively both the perceived slant in this situation and the distortion in perceived slant produced when constant deformation is added to random dot displays depicting moving slanted surfaces. The results confirm that deformation provides a simple account of perceived slant.

INTRODUCTION

It has been long understood that the smooth transformations of the retinal image that accompany observer movement contain useful information about the three-dimensional structure of the world (e.g. Gibson, 1950). Theoretical accounts have identified those aspects of retinal flow that are important in recovering such information and those that are not. For instance, the directional structure of retinal flow merely informs about the direction in which one is heading (e.g. Warren & Hannon, 1988) and the axis about which the eyes (and head) are currently rotating (e.g. Simpson, Graf & Leonard, 1981) because directional structure tells us something about how the world is projected onto the retina but little about what is being projected. Discontinuities in flow speed and/or direction are more informative, because they define where in space object surfaces lie, but they do not tell us much about the properties of those surfaces. On the other hand, the smooth gradients in speed that occur between these discontinuities provide information about 3D surface layout in the form of temporal proximity (Lee, 1980), and surface tilt and slant (Koenderink, 1986). There is considerable evidence that people can make use of these speed gradients in tasks such as estimating temporal proximity (e.g. Lee, 1976), extracting shape (e.g. Rogers & Graham, 1979; Landy, Dosher, Sperling & Perkins, 1991), distinguishing convex from concave surfaces (e.g. Braunstein & Andersen, 1981) and estimating slant (e.g. Braunstein, 1968; Harris, Freeman & Hughes, 1992; Braunstein, Liter & Tittle, 1993).

One particular mathematical analysis of these speed gradients shows that any patch of flow is the sum of a translation ($\text{trans}$), a measure of local size change ($\text{div}$) based upon compressive gradients along the flow, a measure of local rotation ($\text{curl}$) based upon shearing gradients at right angles to the flow, and a measure of local shape change without a change in area ($\text{def}$) based upon both compressive and shearing gradients (Koenderink & van Doorn, 1976). This analysis not only identifies a set of potentially useful primitives for describing general flow, but also specifies the relationships between these primitives and the external world. For example, $\text{def}$ is a vector whose direction and magnitude can be related to surface tilt and slant respectively, which is unaffected by rotations of the observer’s eye or head, and which thus seems ideally suited to the task of estimating surface tilt and slant from general flow (Koenderink, 1986).

Despite the promise of Koenderink’s analysis, it has yet to make much impact upon the psychophysical literature. Here we confirm some of its potential by demonstrating that $\text{def}$ does in fact provide a simple account of the perception of surface slant. We show, firstly, that the combination of $\text{trans}$ and $\text{def}$ is by itself sufficient to produce a powerful and predictable impression of surface slant and, secondly, that the addition of $\text{def}$ to displays depicting slanted virtual surfaces produces predictable distortions of perceived slant.
MODEL

It is well known (e.g., Braunstein, 1968; Harris et al., 1992—see also Fig. 1) that for random dot displays depicting movement of the observer relative to a slanted surface, perceived slant, $s_p$, is linearly related to surface slant, $s_0$, at least over a moderate range of surface slants:

$$s_p = ms_0.$$  \hspace{1cm} (1)

Moreover, for such displays, the def associated with any small patch of the surface is proportional to the tangent of the surface slant. This follows from the fact that the amplitude of def depends on the product of the rate of movement at right angles to the line of sight and the gradient of the surface (Koenderink, 1986). Thus, for a given fixed movement in relation to a particular line of sight:

$$\text{def} = k \tan(s_0)$$

which, upon re-arranging, gives:

$$s_p = \arctan(\text{def}/k).$$ \hspace{1cm} (2)

Substituting equation (2) into (1) predicts a simple relationship between perceived slant and def:

$$s_p = m \arctan(\text{def}/k).$$ \hspace{1cm} (3)

This predicted relationship is confirmed here by presenting observers with displays combining pure trans (to simulate the required relative movement between surface and observer) with varying amounts of pure def and then estimating the resultant perceived slant. Since $m$ is easily estimated in preliminary experiments, the empirical data provide an estimate of the single parameter $k$.

With the parameters $m$ and $k$ now fixed, the relationship can be further investigated by adding known amounts of pure def to displays depicting movement relative to a slanted surface. This simply adds a constant, $d$, to equation (3):

$$s_p = m \arctan[(\text{def} + d)/k].$$ \hspace{1cm} (4)

Our results show that resultant slant estimates in this circumstance are well explained by equation (4).

STIMULI

The surface component, S, was initially produced by polar projection of the position of each dot onto a virtual, flat surface, located 285 cm from the observer's eye, and already tilted and slanted by the required amount $s_0$. This technique removes gross texture density and shape cues to slant from the display (although at the cost of introducing additional cues to flatness). The resulting virtual surface moved sinusoidally at 1 Hz along a horizontal trajectory in the frontoparallel plane through a virtual peak-to-peak distance of 12 cm.

The affine component, A, was determined by the general affine transformation matrix:

$$A = \begin{pmatrix} +1 & 0 \\ 0 & +1 \end{pmatrix} + \text{curl} \begin{pmatrix} 0 & -1 \\ +1 & 0 \end{pmatrix} + \text{def} \begin{pmatrix} -1 & 0 \\ 0 & +1 \end{pmatrix} R^{-1}$$

where $R$ is the standard rotation matrix:

$$R = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}$$

and $\theta$ is the angle of the axis of compression, or orientation, of def, measured anticlockwise from horizontal.

The amplitudes of the individual affine coefficients (div, curl and def) were also modulated sinusoidally over time, in phase with that of the surface component, S. Thus, for example, $\text{def} = \text{def}_{\text{max}} \sin(2\pi t)$ and all values quoted here refer to amplitude, $\text{def}_{\text{max}}$.

PROCEDURE

Two of the authors (TCAF & TSM) undertook the experiments. Each session consisted of 25 trials for TCAF and 32 trials for TSM. On each trial the
observer's task was to match the perceived slant of the display using a pictorial wire-frame cube that could be rotated on the PC screen positioned alongside the display. This stimulus is naturally interpreted as an object in a specific pose and thus encourages a 3D match. In practice, we have found that it produces identical results to the more traditional manipulation of a real planar surface (e.g. Braunstein, 1968) but is more convenient for both subject and experimenter. Both displays remained visible until the observer was satisfied with the match. All experiments were conducted in a darkened room. The observers viewed the display monocularly, using their preferred eye, and were not instructed to adopt any particular fixation strategy. Both observers undertook several practice sessions before any data were collected.

In Experiment 1, the parameter $m$ was estimated (equation 1) by setting the affine component to zero and, on each trial, randomly selecting a slant for the surface component from the range $-60 < s_0 < 60$. TCAF undertook one session with surfaces tilted about a horizontal axis, which we define as tilt = 90°, and one session with surfaces tilted about a vertical axis (tilt = 0°). TSM completed two sessions for each tilt direction.

In Experiment 2, the relationship between perceived slant, $s_v$, and def (equation 3) was investigated by setting the slant of the surface component to zero (thus producing pure trans) and, on each trial, randomly selecting a value of $\text{def}_{\text{max}}$ from the range $-0.05 < \text{def}_{\text{max}} < +0.05$. Each observer undertook two sessions for both orientations of def.

In Experiment 3, the effect of adding pure def to a display depicting a moving, slanted surface (equation 4) was investigated by setting the magnitude $\text{def}_{\text{max}}$ to $\pm 0.00624, \pm 0.0125, \pm 0.025,$ or $\pm 0.05$ and, on each trial, randomly selecting a slant for the surface component from the range $-60° < s_0 < 60°$. Each observer experienced the 16 conditions resulting from the individual combinations of 4 amplitudes, 2 polarities and 2 orientations of def. For TCAF, one value of $\text{def}_{\text{max}}$ was randomly chosen and fixed for each session of 25 trials. For TSM, the value of $\text{def}_{\text{max}}$ was randomised from trial to trial within each session of 32 trials.

RESULTS

Figure 1 plots the results of Experiment 1 and supports the previously reported finding that the relationship between surface slant and perceived slant is well described by a straight line over the range tested here ($60° < s_0 < 60°$). For simplicity we estimate only the slope ($m$) of this linear relationship, ignoring any intercept. (In practice, when the intercept is included in the model, it makes negligible difference to the final outcome.) The slope estimates of around 0.5 are similar to those we have previously reported under similar viewing conditions.
conditions and we have previously discussed how such systematic underestimation of perceived slant may be related to the flatness cues provided by the display (Harris et al., 1992).

In Experiment 2, preliminary sessions in which either curl or div, rather than def, were added to the translating display showed that neither of these components produced any compelling impression of surface slant. However, the addition of pure def produced an immediate and compelling impression of slant for both observers, demonstrating that def provides a sufficient basis for perceiving slant. Figure 2 confirms, for both observers, the relationship between perceived slant and def predicted by equation (3). The curves were derived using the individual values of m found for each observer in Experiment 1 and the parameter k was fitted using a least squares technique.

As an alternative to equation (1), we can represent the relationship between perceived and physical slant as \(\tan(s_p) = m \tan(s)\), which leads to the alternative model \(s_p = \arctan(m \text{def}/k)\).* Unlike equation (3), this allows perceived slants of up to 90 degrees. We have not yet systematically investigated performance at high slants and, with the current data, the two models are not reliably distinguishable. However, one initial reason for favouring equation (3) is our consistent impression that perceived slant does saturate below 90 degrees under the conditions reported here.

Experiment 3 investigated the effects of adding the appropriate orientation of pure def to a display depicting a surface slanted about a tilt axis of either 0° or 90°. Figures 3–6 show that, for both observers, the resulting distortion of perceived slant is well explained by equation (4) when either positive (closed symbols) or negative (open symbols) polarities of def are added to the displays. The simple model, with only two parameters, both of which are fixed from the previous experiments, captures both the reduction in overall slope and the change in offset as the magnitude of additional def is increased. Any occasional discrepancies that exist for one subject do not appear to be mirrored by the other.

**DISCUSSION**

The model appears to capture quite closely both the perceived slant induced by combining pure trans and def and the distortion in perceived slant produced by adding pure def to displays depicting moving, slanted surfaces. Koenderink’s formal analysis (Koenderink & van Doorn, 1976; Koenderink, 1986) identifies div, curl, def.
and trans as potentially useful primitives for describing complex natural flow patterns, in the same way and for the same reasons that sinewave gratings can be used to describe complex, natural luminance patterns. The analysis further specifies the relationships between these primitives and such physical variables as tilt and slant. The work described here confirms some of the empirical promise of this theoretical work by demonstrating that the descriptive primitive def can be used to predict the relationship between these physical variables and their perceptual correlates.

One of the most intuitively appealing aspects of an approach based on def is that it offers an explicitly 2D analysis of what is generally a 2D problem. It is true that retinal flow can, in some circumstances, be reduced to 1D: for example, horizontal motion relative to a
vertically tilted surface produces a 1D vertical shearing gradient, while the same motion relative to a horizontally tilted surface produces a 1D horizontal compression gradient. But analyses based only on such 1D gradients (e.g. Braunstein et al., 1993) provide no immediate account of our finding that def produces a compelling impression of slant whereas curl and div do not. For, as shown in Fig. 7, these stimuli contain the appropriate 1D gradients and differ only in the sign of their orthogonal component. Interestingly, an exactly analogous situation arises in stereopsis, because the underlying geometry is the same, and one can think of def in terms of disparity composed of horizontal and vertical shearing gradients. Howard and Kaneko (1994) have demonstrated that both of these 1D gradients contribute to perceived stereoscopic slant and we have recently shown the same for retinal flow with both shearing and compressive gradients. (Meese, Harris & Freeman, 1995).

Finally, despite its obvious promise, our model needs refining in at least two ways. Firstly, though it makes the correct qualitative predictions, the fit to the data is far from perfect. One reason for this may be the simplicity of our initial approach, which assumes that our stimuli are interpreted entirely within the context of translational motion of the observer or the surface. A more sophisticated model will need to consider both translational and rotational motion (Braunstein et al., 1993). Secondly, a measure of def alone is not sufficient for the visual system to recover surface slant but must first be related to some measure of trans, because a given value of def might be produced by rapid movement relative to a slightly slanted surface, or by slower movement relative to a more slanted surface. We would thus expect our parameter k to vary systematically with trans. We are currently investigating both of these important refinements.

REFERENCES


Acknowledgement—This work was supported by SERC Grant GR/H/52986 to MGH
Section 4

Computational Vision