

Recovery of surface pose from texture orientation statistics under perspective projection

Paul A. Warren · Pascal Mamassian

Received: 22 April 2009 / Accepted: 31 March 2010 / Published online: 28 April 2010
© Springer-Verlag 2010

Abstract In a seminal paper, [Witkin \(1981\)](#) derived a model of surface slant and tilt recovery based on the statistics of the orientations of texture elements (texels) on a planar surface. This model made use of basic mathematical properties of probability distributions to formulate a posterior distribution on slant and tilt given a set of image orientations under orthographic projection. One problem with the Witkin model was that it produced a posterior distribution with multiple maxima, reflecting the inherent ambiguity in scene reconstruction under orthographic projection. In the present article, we extend Witkin's method to incorporate the effects of perspective projection. An identical approach is used to that of Witkin; however, the model now reflects the effects of perspective projection on texel orientation. Performance of the new model is compared against that of Witkin's model in a basic surface pose recovery task using both a *maximum a posteriori* (MAP) decision rule and a rule based on the expected value of the posterior distribution. The resultant posterior of the new model is shown to have only one maximum and thereby the ambiguity in scene interpretation is resolved. Furthermore, the model performs better than Witkin's model using both MAP and expected value decision rules. The results are discussed in the context of human slant estimation.

Electronic supplementary material The online version of this article (doi:[10.1007/s00422-010-0389-3](https://doi.org/10.1007/s00422-010-0389-3)) contains supplementary material, which is available to authorized users.

P. A. Warren (✉)
School of Psychological Sciences,
Manchester University, Manchester, M13 9PL, UK
e-mail: paul.warren@manchester.ac.uk

P. Mamassian
Laboratoire Psychologie de la Perception (CNRS UMR 8158),
Université Paris Descartes, 45 rue des Saints-Pères, 75006 Paris, France

Keywords Shape from texture · Computational vision · Human vision

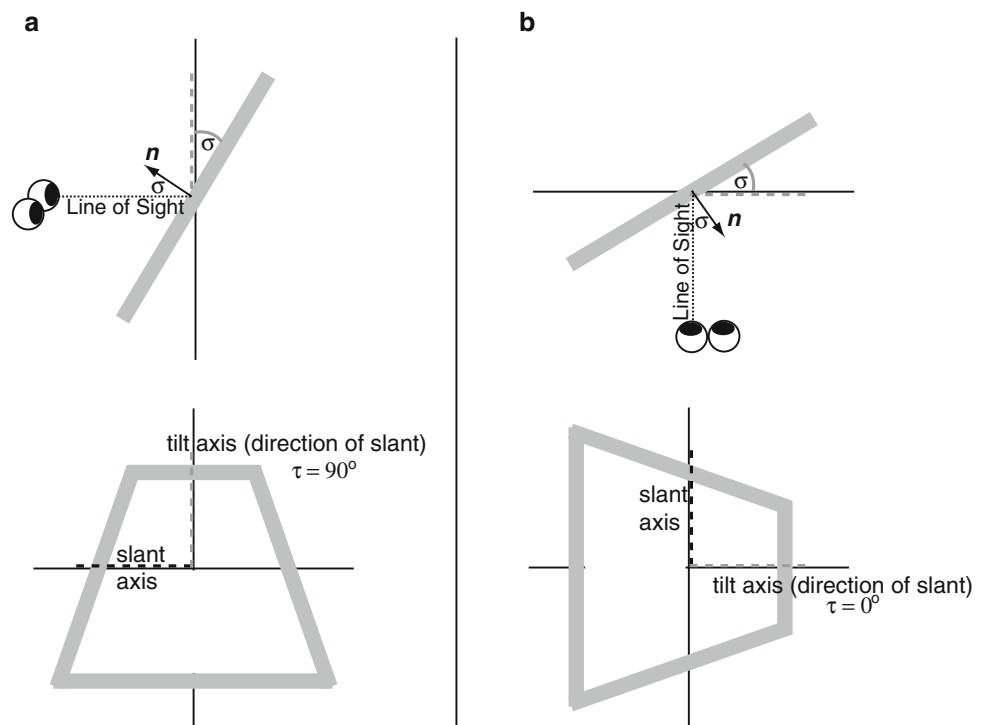
1 Introduction

A fundamental problem faced by any (animal or artificial) visual system is to recover the 3D structure of the scene from the 2D information available in the image. One particular problem of interest involves estimation of the shape of objects in the scene. This is equivalent to the recovery of the local pose (i.e., the direction of the surface normal) of small planar regions on the object surface. Undertaking such a process over many such regions allows the visual system to build up a global estimate of object shape.

In theory, local surface normals can be recovered from a number of different information sources including surface shading (e.g., see [Horn and Brooks 1989](#)) and the properties of texture markings on the surface (e.g., see [Witkin 1981](#); [Gårding 1993a](#); [Stone 1993](#)). Furthermore, the ability of human observers to recover surface shape and orientation from shading and texture has been demonstrated ([Mingolla and Todd 1986](#); [Koenderink et al. 1992](#); [Rosenholtz and Malik 1997](#)) and contrasted with performance of ideal observers ([Blake and Marinos 1990](#); [Blake et al. 1993](#); [Buckley et al. 1995](#); [Knill 1998](#); [Rosenholtz and Malik 1997](#)).

In the present article, we focus on the problem of recovering planar surface pose, characterised by the slant and tilt of the surface normal (see [Fig. 1](#)), from texture information. In particular, we investigate the recovery of surface pose from the image projection of discrete markings or texture elements (texels) on the surface. Several algorithms have been developed which can recover surface shape without the need for texels, e.g., those based on deformations of oriented spectral content such as [Li and Zaidi \(2000\)](#), deformations in

Fig. 1 Illustration of surface pose parameters—slant, σ , and tilt, τ . Note that slant as defined in the text is equivalent to the more traditional definition of slant as the angle between the line of sight and the normal to the surface. Planes are shown slanted about an **a** horizontal and **b** vertical axis. In both **a** and **b**, the top figure shows a view of the plane from the side (**a**) or from the top (**b**). The bottom figure shows a schematic illustration from the point of view of the observer and indicates the tilt and slant axes



affine structure, e.g., [Malik and Rosenholtz \(1997\)](#) or geometric distortions of closed contours, e.g., [Brady and Yuille \(1984\)](#). However, there is evidence that regular local texture elements form an important part of human slant recovery ([Velisavljevic and Elder 2006](#)). Furthermore, when texels are projected into the image a number of simple but powerful effects on the structure of the texture occur, which provide useful information about surface pose. These effects can be summarised as:

- (i) Size—as the slant of the plane is increases, the projected texels become relatively smaller and larger in the portions of the plane which are located, respectively, farther from and nearer to the camera or eye.
- (ii) Density—as the slant of the plane is increased, the projected texels become relatively more and less dense in the portions of the plane which are located, respectively, farther from and nearer to the camera or eye.
- (iii) Compression/foreshortening—as the slant of the plane is increased the aspect ratio of texels varies systematically so that the extent of the texel in the direction of the tilt axis is compressed relative to that in the orthogonal direction.

In this study, we are particularly interested in the effects of compression and how they can be used to recover the pose (slant and tilt) of a planar surface. Both size and density cues can be used to recover surface pose, however, they rely upon the rate of change of texture information across the scene,

and as such they rely upon gradients in texture information ([Gibson 1950](#)). Furthermore, for small surfaces the size and density cues are of limited use. The effect of foreshortening is different from size and density cues in that information about surface pose can be recovered from any single texel on the surface and therefore does not rely upon a gradient of information across the scene or the size of the surface.

One consequence of compression is that under the assumption that all texel orientations are equally likely on the surface (the isotropy assumption); the projected orientations in the image approach the orientation of the axis of slant. This tendency is illustrated in [Fig. 2](#) for a plane covered in texels whose orientations are selected randomly from a uniform distribution. As the slant about a horizontal axis increases, the distribution of orientations in the image becomes more peaked around the horizontal direction.

[Witkin \(1981\)](#) provided an elegant algorithm which, under the isotropy assumption, takes advantage of the systematic relationship between the slant and tilt of the plane and the distribution of discrete texel orientations in the image to recover surface orientation. Witkin's approach rests upon the idea of having a statistical model of the texture on the surface. Recovering the surface pose then becomes a problem of statistical inference; slant (σ) and tilt (τ) are parameters to be estimated given observed data in the image. Witkin provides a statistical estimator (Eq. 1) for (σ, τ) , based on the posterior distribution $p(\sigma, \tau | \theta')$, describing the probability that the surface has slant σ and tilt τ given the n -vector of image texel orientations $\theta' = \{\theta'_1, \theta'_2, \dots, \theta'_n\}$:

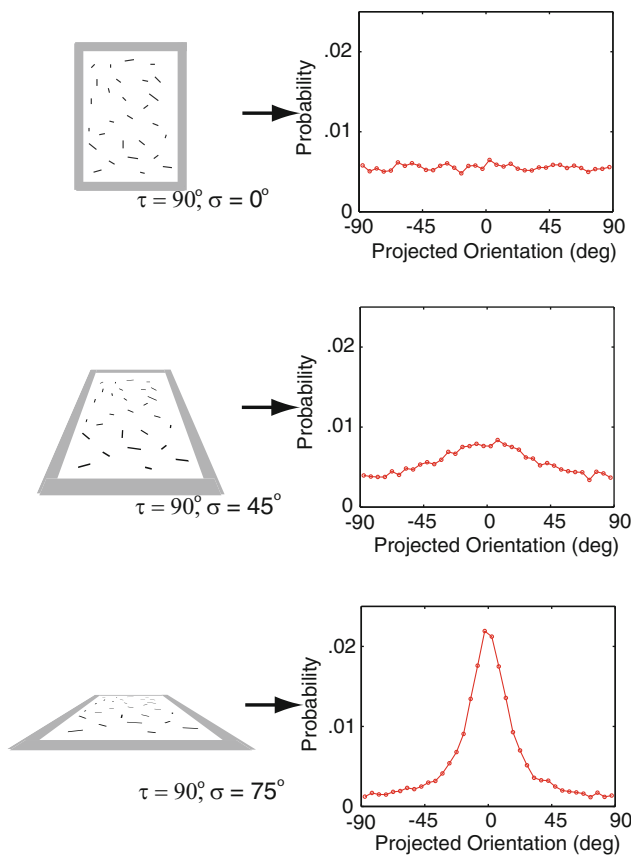


Fig. 2 Schematic illustration of effects of foreshortening on the distribution of texel orientations in the image. The *top* row shows a front-to-parallel surface on which are placed a number of texels at random orientations as illustrated in the approximately uniform orientation distribution. The *middle* row shows the effect of slanting this surface about a horizontal axis. A peak is formed in the distribution of texel orientations at a location which characterises the axis of tilt. The *bottom* row demonstrates that as the slant is increased the peak in the distribution becomes more concentrated around a direction which depends on the tilt

$$p(\sigma, \tau | \theta') = \prod_{i=1}^n \frac{\pi^{-2} \sin \sigma \cos \sigma}{\cos^2(\theta'_i - \tau) + \sin^2(\theta'_i - \tau) \cos^2 \sigma} \quad (1)$$

This method is by no means the only approach to recovery of surface pose from texture information in the image. Kanatani (1984) proposed an alternative continuous contour-based approach based on counting intersections of contours with straight lines in the image plane. Brady and Yuille (1984) applied another approach for closed contours on a surface, suggesting that surface pose could be recovered by maximizing compactness (a metric related to the enclosed area and perimeter length of the contour) in the back projection onto the surface. Furthermore, Blake and Marinos (1990) and Gårding (1993a) have developed iterative methods to recover approximations to the maximum likelihood estimates of slant and tilt as described by Witkin (1981).

One problem with the Witkin algorithm and the majority of alternative approaches to recover surface pose, is that the geometric model is based upon orthographic (parallel) projection so that orientations in the image are independent of texel location on the surface. Under this projection model, the image data obtained for a surface with pose parameters (σ, τ) is identical to that when the plane has pose $(\sigma, \tau + \pi)$ (see eq. 1). As a consequence when such models are shown a set of randomly oriented texels, then simply by chance (i.e., depending on the set of texels generated) an error could arise of around 180° in the recovered tilt.

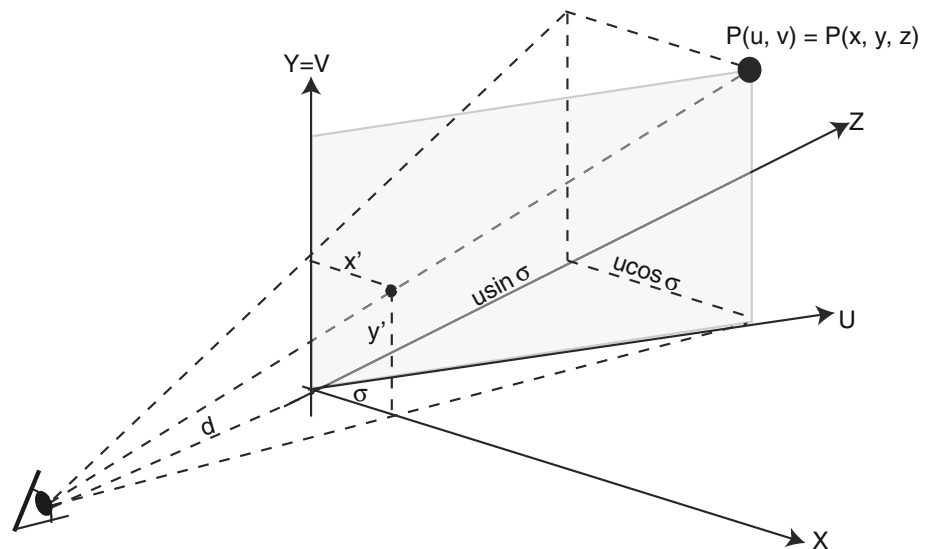
The perspective projection case has been considered by Gårding (1993b). An algorithm (Weak Isotropy Surface Perception or WISP) was proposed incorporating a perspective projection model and using standard results from circular statistics. The algorithm relies upon an initial assumption of strong isotropy which is then relaxed to assume only weak isotropy (isotropy of texture on the average) in an iterative procedure. However, this approach still recovers two rival interpretations of the surface pose which are then manually selected. The reason for this ambiguity is that under the WISP model, the texels are all treated as if they occur at the centre of the image. Thus, in spite of perspective projection information being present in the shape of the texel in the image, the position of those texels is not taken into account. Consequently, this model suffers from a more complex analogue of the ambiguity arising under orthographic projection.

In the present article, we extend Witkin's original algorithm directly to take into account a perspective projection model. We derive a generalization of (Eq. 1) for perspective projection which provides an estimator for (σ, τ) based on $p(\sigma, \tau | \theta', \mathbf{u}, \mathbf{v})$, the probability that the surface has slant σ and tilt τ given the n -vector of image orientations θ' and the 2D locations of the line segments on the plane (\mathbf{u}, \mathbf{v}) . Throughout this article, we will use the coordinate system described in Fig. 3.

The model is compared to the orthographic projection method of Witkin (1981) using two different decision rules. When either a maximum a posteriori (MAP) or expected value (EXP) decision rule are imposed upon the posterior, the estimate is seen to have a single value (i.e., the tilt ambiguity is removed). The reason for investigating performance using the EXP decision rule is that it has been suggested that human performance in tasks of this kind might be better represented by sampling on a trial by trial basis from the posterior distribution (Mamassian and Landy 1998). Consequently, the expected value of the posterior might be more appropriate for representation of human slant estimation performance.

The results of our model simulations are discussed in the context of human slant estimation. We also discuss the similarities between our model and some human data collected in a slant estimation experiment (presented in the supplementary materials). These experiments were designed to explore

Fig. 3 Coordinate system used throughout this study



the common pattern of slant underestimation found in previous studies (e.g., Gibson 1950; Gruber and Clark 1956; Braunstein 1968) and to see if this is consistent with the results of our model. Furthermore, we investigated whether human participants are sensitive to texel orientation information under perspective projection.

2 Methods: deriving the estimator

As in Witkin (1981) we will split the derivation into sections corresponding to the geometric and statistical models before finally characterising the estimator for slant and tilt. At each stage we will demonstrate how the present model is a generalisation of the Witkin model.

As noted in the introduction, the Witkin (1981) algorithm is based upon an orthographic projection model. Consequently, the location of line elements on the plane is not factored into the estimator. In the first section we derive some basic results for perspective projection of surface texels to the image plane.

2.1 The projection model—Case 1: $\tau = 0$

We begin by deriving the function describing how locations and angles on the surface are transformed under perspective projection for the zero tilt case (i.e., rotation about the Y-axis; see Fig. 1).

Let the triple $\{X, Y, Z\}$ define an orthogonal, reference coordinate frame with origin at a distance d from the observer (Fig. 3). Let P be a point on a vertical plane passing through the origin of this frame and slanted about the Y-axis by an angle σ . Let $\{U, V\}$ define the coordinate system attached to this plane. If we define the image plane to be the X-Y plane then by similar triangles we have:

$$x = \frac{ud \cos \sigma}{d + u \sin \sigma}, \quad y = \frac{vd}{d + u \sin \sigma} \quad (2)$$

where d is the viewing distance. This equation provides the coordinates (x, y) of point P in the image plane under perspective projection.

Now let $(u_1, v_1)^T$ and $(u_2, v_2)^T$ be two such points on the surface and let the line between them define angle θ with the U-axis. By Eq. 2, the projections of these points in the image are given by

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} \frac{u_1 d \cos \sigma}{d + u_1 \sin \sigma} \\ \frac{v_1 d}{d + u_1 \sin \sigma} \end{pmatrix}, \quad \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} \frac{u_2 d \cos \sigma}{d + u_2 \sin \sigma} \\ \frac{v_2 d}{d + u_2 \sin \sigma} \end{pmatrix}.$$

Thus, relative to the X-axis, θ' , the projection of angle θ in the image, is given by:

$$\tan \theta' = \frac{y_2 - y_1}{x_2 - x_1} = \frac{d(v_2 - v_1) + \sin \sigma (v_2 u_1 - v_1 u_2)}{d \cos \sigma (u_2 - u_1)} \quad (3)$$

However, since the points on the surface define a straight line the following relationships hold.

$$u_2 = u_1 + \alpha \cos \theta, \quad v_2 = v_1 + \alpha \sin \theta, \quad \text{so that:}$$

$$\begin{cases} u_2 - u_1 = \alpha \cos \theta \\ v_2 - v_1 = \alpha \sin \theta \\ v_2 u_1 - v_1 u_2 = \alpha (u_1 \sin \theta - v_1 \cos \theta). \end{cases}$$

Substituting these into Eq. 3, we obtain θ' from

$$\begin{aligned} \tan \theta' &= \frac{\tan \theta + \frac{1}{d} \sin \sigma (u_1 \tan \theta - v_1)}{\frac{\tan \theta (1 + \frac{u_1}{d} \sin \sigma) - v_1/d \sin \sigma}{\cos \sigma}} \\ &= \frac{\tan \theta + \frac{1}{d} \sin \sigma (u_1 \tan \theta - v_1)}{\cos \sigma} \end{aligned} \quad (4)$$

Note that for the cases in which (i) d is large or (ii) the line is close to the origin (u and v are small) Eq. 4 is approximated by

$$\tan \theta' = \frac{\tan \theta}{\cos \sigma},$$

which is the basic orthographic projection model from Witkin (1981).

2.2 The projection model—Case 2: arbitrary τ

To introduce the effect of surface tilt on projected image orientation we now pick arbitrary coordinate axes for the image plane (namely X_τ and Y_τ) with X_τ in the direction of the tilt axis which is inclined at an angle of τ relative to the X -axis. We also define θ' as the angle between the X -axis and a projected line orientation, together with θ'_τ as the angle between the X_τ -axis and the projected line orientation. Then, by our definition we know that

$$\theta'_\tau = \theta' - \tau$$

Also, using (4) we can calculate the angle θ'_τ as

$$\theta'_\tau = \tan^{-1} \left[\frac{\tan \theta_\tau (1 + \frac{u}{d} \sin \sigma) - \frac{v}{d} \sin \sigma}{\cos \sigma} \right]$$

where θ_τ is the orientation of the line element on the surface prior to projection relative to the projection of the X_τ -axis onto the surface. Consequently, we obtain (Eq. 5) as an expression for the projected orientation of a line segment on the surface as a function of the surface's slant and tilt and the line's orientation and position on the surface.

$$\begin{aligned} \theta' &= f(\theta_\tau; \sigma, \tau, u, v) \\ &= \tan^{-1} \left[\frac{\tan \theta_\tau (1 + \frac{u}{d} \sin \sigma) - \frac{v}{d} \sin \sigma}{\cos \sigma} \right] + \tau \end{aligned} \quad (5)$$

Note that for the cases in which (i) d is large or (ii) the line is close to the origin (u and v are small) Eq. 5 is approximated by:

$$\theta' = \tan^{-1} \left[\frac{\tan \theta_\tau}{\cos \sigma} \right] + \tau$$

which is Eq. 1 from Witkin (1981).

2.3 Statistical model (isotropy and independence)

We use the statistical model of Witkin (1981). Each of the texels on the surface, with orientation θ_τ , projects to a line in the image, with orientation θ' . The distributions of orientations on the surface and across the whole image plane are related by the geometric transformation in Eq. 5.

By assuming a particular form for the distribution of orientations on the surface it is possible to quantify the likelihood that an observed image orientation occurred for each possible slant and tilt pair. In Witkin (1981) it is assumed that the distribution of orientations on the surface is uniform (the isotropy assumption).

In addition, assuming that the line orientations on the surface are independent from one another (the independence assumption), it is possible to quantify the combined likelihood that the set of observed image orientations occurred for each possible slant and tilt pair.

Since Witkin's method was based upon an orthographic projection model there was no dependence on texel orientation on the plane. Under the perspective projection model used in the present study we make the additional assumption that the location of texels is uniformly distributed on the surface (the homogeneity assumption).

A final assumption is that all surface normals describing the orientation of the plane are equally likely. Consequently, we do not implement an explicit prior belief that floor or ceiling planes are more likely in the world. This enables us to assess the performance of our estimator without bias due to prior assumptions.

2.4 Deriving the estimator for surface slant and tilt

2.4.1 Step 1: a single texel

We wish, first of all, to characterise the likelihood $L(\sigma, \tau; \theta', u, v)$ that each slant and tilt pair generated an observed single image orientation θ' from the surface location (u, v) , under the geometric and statistical models introduced above. It should be noted at this point that the likelihood appears to be a mixed function of the post-projection image orientation and the pre-projection position on the surface (u, v) . Note, however, that u and v are simply parameters in the likelihood and not the variables of interest. In fact, by simply inverting Eq. 2, the position of the texel in the image could also be substituted in Eq. 6, so that all parameters are expressed post-projection. We choose not to do this solely for the sake of parsimony. This likelihood function can be determined by treating the conditional probability distribution $p(\theta'|\sigma, \tau; u, v)$ as a function of slant and tilt—note that the semi-colon indicates that u and v are simply parameters. To derive this quantity we will use the isotropy assumption which constrains the distribution of θ_τ to be

$$p(\theta_\tau) = \frac{1}{\pi}.$$

We will also use the geometric model expressed in Eq. 5 which determines θ' as a function of θ_τ (and vice versa) with σ, τ, u and v as parameters. Finally, we use the standard relation for determining the probability distribution function (p.d.f.) of a function, ξ , of a random variable, x , with known distribution:

$$\text{pdf}(\xi(x)) = \text{pdf}(x) \frac{dx}{d\xi(x)}$$

Replacing x with θ_τ and ξ with f (the function from Eq. 5), we can derive the required p.d.f., $p(\theta'|\sigma, \tau; u, v)$, of

a projected orientation given the slant and tilt of the surface and the surface position of the texel:

$$p(\theta'|\sigma, \tau; u, v) = p(\theta_\tau) \frac{d\theta_\tau}{d\theta'} = \frac{1}{\pi} \frac{d\theta_\tau}{d\theta'}.$$

In order to carry out the differentiation we require θ_τ , expressed as a function of θ' , which is obtained by inverting Eq. 5:

$$\begin{aligned} \theta_\tau &= g(\theta'; \sigma, \tau, u, v) \\ &= \tan^{-1} \left[\frac{\tan(\theta' - \tau) \cos \sigma + v/d \sin \sigma}{(1 + u/d \sin \sigma)} \right]. \end{aligned}$$

To calculate the derivative we set the term inside the square bracket equal to γ and use the chain rule:

$$\frac{d\theta_\tau}{d\theta'} = \frac{d\theta_\tau}{d\gamma} \frac{d\gamma}{d\theta'} = \frac{1}{1 + \gamma^2} \frac{d\gamma}{d\theta'}.$$

Using the quotient rule to evaluate the remaining derivative and simplifying we obtain:

$$\frac{d\theta_\tau}{d\theta'} = \frac{\sec^2(\theta' - \tau) \cos \sigma (1 + u/d \sin \sigma)}{(1 + u/d \sin \sigma)^2 + (\tan(\theta' - \tau) \cos \sigma + v/d \sin \sigma)^2}.$$

As a consequence we obtain the required likelihood function:

$$\begin{aligned} L(\sigma, \tau; \theta', u, v) &= p(\theta'|\sigma, \tau; u, v) = \frac{1}{\pi} \frac{d\theta_\tau}{d\theta'} \\ &= \frac{\pi^{-1} \sec^2(\theta' - \tau) \cos \sigma (1 + u/d \sin \sigma)}{(1 + u/d \sin \sigma)^2 + (\tan(\theta' - \tau) \cos \sigma + v/d \sin \sigma)^2}. \end{aligned} \quad (6)$$

Note that for the cases in which (i) d is large or (ii) the line is close to the origin (u and v are small) Eq. 6 is approximated by the Witkin (1981) likelihood function.

2.4.2 Step 2: multiple independent texels

To obtain a joint density function for n image texel orientations defined in the vector $\theta' = \{\theta'_1, \theta'_2, \dots, \theta'_n\}$, we use the assumption of independence stated in the statistical model section. Let $\mathbf{u} = \{u_1, u_2, \dots, u_n\}$ and $\mathbf{v} = \{v_1, v_2, \dots, v_n\}$ be vectors containing the position of the texels then

$$L(\sigma, \tau; \theta', \mathbf{u}, \mathbf{v}) = \prod_{i=1}^n L(\sigma, \tau; \theta'_i, u_i, v_i). \quad (7)$$

It can be seen from Eqs. 6 and 7 that for the cases in which (i) d is large or (ii) the texels were close to the origin before projection (i.e., u_i and v_i are small) Eq. 7 reduces to Eq. 1.

2.4.3 Step 3: application of Bayes' rule

From Bayes' rule, we know that the posterior density function, $p(\sigma, \tau|\theta', \mathbf{u}, \mathbf{v})$, allowing estimation of the slant and tilt

parameters, can be obtained as:

$$p(\sigma, \tau|\theta'; \mathbf{u}, \mathbf{v}) = \frac{p(\sigma, \tau) p(\theta'|\sigma, \tau; \mathbf{u}, \mathbf{v})}{\iint p(\sigma, \tau) p(\theta'|\sigma, \tau; \mathbf{u}, \mathbf{v}) d\sigma d\tau}$$

where the integration is performed over the σ and τ ranges and $p(\sigma, \tau)$ is the joint prior probability of each slant, tilt pair.

To obtain $p(\sigma, \tau)$ we use the assumption that all surface normals are equally likely (see statistical model section). It can be shown (e.g., see Witkin 1981) that $p(\sigma, \tau)$ is then given by:

$$p(\sigma, \tau) = \frac{\sin \sigma}{\pi}.$$

Consequently, the relative likelihood (i.e., the posterior distribution before normalisation) is given by

$$\begin{aligned} p(\sigma, \tau) p(\theta'|\sigma, \tau; \mathbf{u}, \mathbf{v}) &= \frac{\sin \sigma}{\pi} \prod_{i=1}^n \\ &\frac{\pi^{-1} \sec^2(\theta'_i - \tau) \cos \sigma (1 + u_i/d \sin \sigma)}{(1 + u_i/d \sin \sigma)^2 + (\tan(\theta'_i - \tau) \cos \sigma + v_i/d \sin \sigma)^2}. \end{aligned} \quad (8)$$

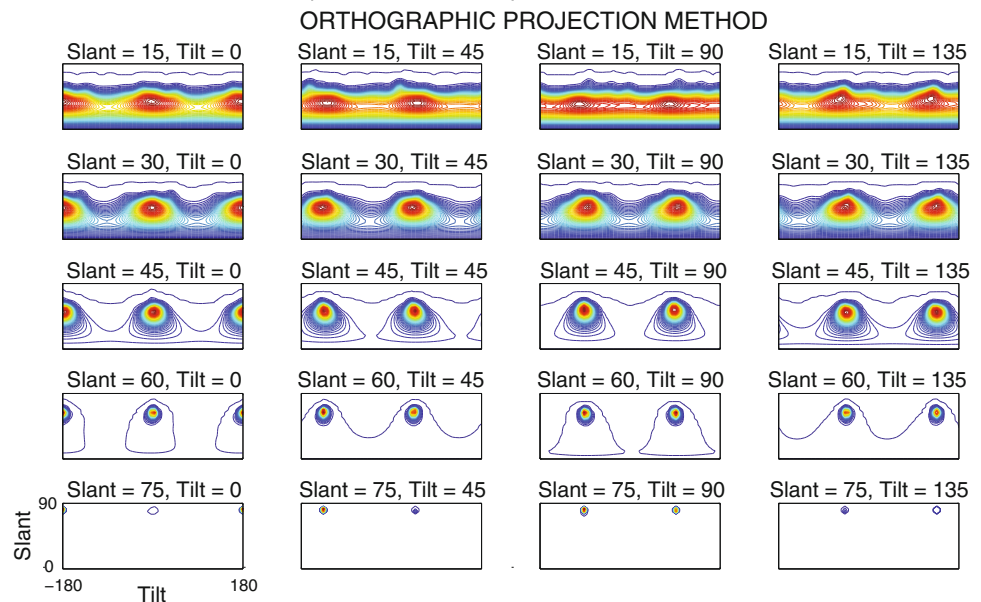
Note, once again, that for the cases in which (i) d is large or (ii) u and v are small, Eq. 8 is approximated by Eq. 3 from Witkin (1981). Normalising this expression with respect to its integral over the slant and tilt ranges will yield the required posterior distribution.

3 Simulations

In this section, we present the results of simulations which compare the performance of the present method (referred to as the perspective projection method) and that from Witkin (1981) (referred to as the orthographic projection method). The models will 'view' images containing a number of oriented line segments obtained under perspective projection from a planar surface. One hundred trials were conducted for each pair of 5 slant (15°, 30°, 45°, 60°, 75°) and 4 tilt (0°, 45°, 90°, 135°) values defining the surface normal direction. For the sake of clarity, these are referred to as 'world' slants and tilts (σ_w, τ_w), in contrast to the 'estimated' slants and tilts (σ_e, τ_e) obtained from the estimator.

On the surface, the lines were randomly positioned (following a 2D uniform distribution) and orientated (following a uniform distribution). Note that for the simulations undertaken, for both the orthographic and the perspective model, the texels were projected into the image using perspective projection. Consequently, the two models have access to the same data, but for the perspective model the estimator for surface pose uses a p.d.f. for texel orientation in the image which takes location into account. The competing methods will return estimates of surface slant and tilt over the ranges

Fig. 4 Contour plots derived from posterior distribution for orthographic projection method model of Witkin (1981) for a range of world slants and tilts (each subplot corresponds to a different world slant and tilt pair). Note that there are always two approximately equal peaks in the distribution reflecting the inherent ambiguity in this model between planes which have tilts differing by 180° . Different random samples of texels on the surface will lead to one peak being favoured over the other at random



$[0, \pi/2]$ and $[-\pi, \pi]$, respectively. These ranges define a hemisphere of possible surface normals.

Simulations are conducted for a scenario in which the viewing distance is 57 cm and the observer views a square plane (with side length 2.24 m) through a circular aperture with a diameter of 20° of visual angle. On the plane, 10,000 line segments are positioned at random (heterogeneity assumption) with random orientation (isotropy assumption). Texel density on the surface is therefore approximately 0.2 texels/cm^2 . Elements falling outside the 20° field of view after projection were not ‘seen’ by the model. The size of this plane was chosen so that, for this viewing distance, even in the largest slant condition simulated, the projected texels filled the aperture (i.e., the top edge of the plane never entered the aperture).

One final aspect of the model to be specified is the decision rule employed to convert the posterior distribution into an estimate of slant and tilt. Witkin (1981) used the MAP solution and accordingly we will present the same solution in what follows. However, we will also show results using an alternative decision rule (referred to as EXP) in which the spherical expectation is calculated over the slant and tilt hemisphere (e.g., see Mardia and Jupp 1999). This decision rule was also used since there is some suggestion in the literature that it would better describe the behaviour of a human decision maker who samples from the posterior distribution (see Mamassian and Landy 1998).

3.1 Results—Orthographic projection method

Figure 4 shows contour plots for the set of posterior distributions averaged over the 100 trials, for each of the (σ_w, τ_w)

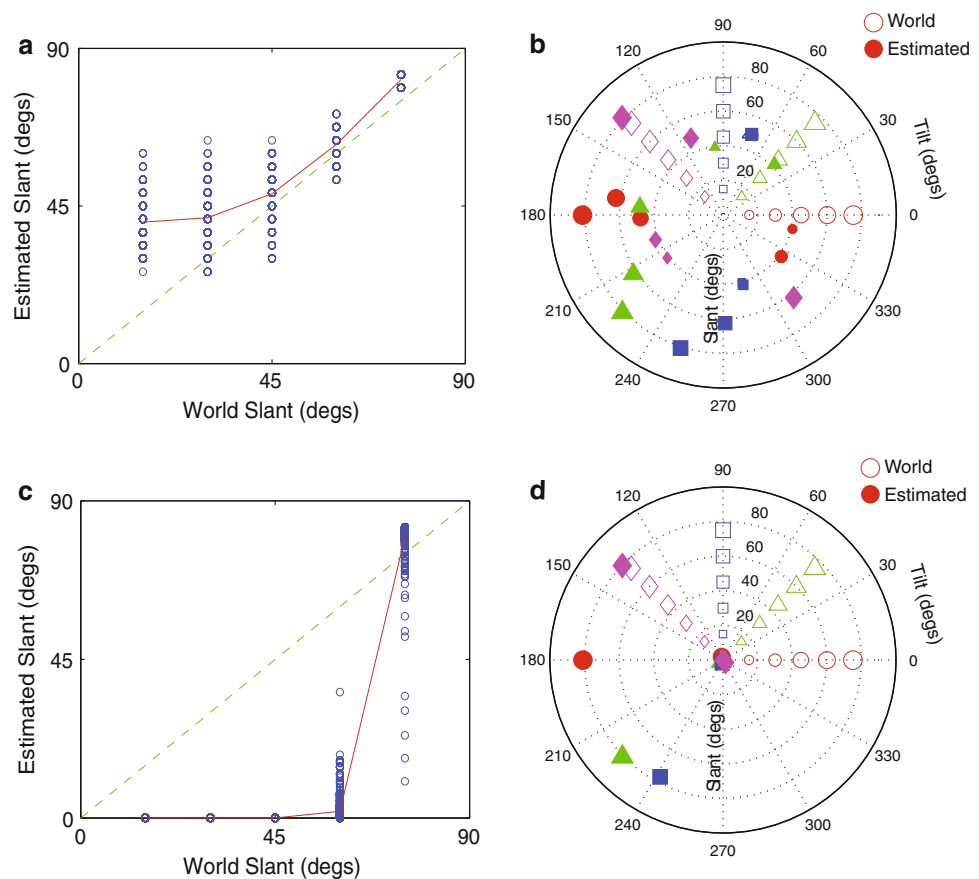
conditions. Care should be taken when examining these plots since we are representing a circular quantity (tilt) on a linear scale, consequently the left and right ends of the tilt axis should be regarded as equivalent. First note that the likelihood becomes more peaked as the slant increases and consequently the variability in the slant and tilt estimate decreases with the real surface slant. Note also that there appear to be two peaks for each posterior. This result reflects the fact that under an assumption of orthographic projection image data obtained for a plane in the (σ_w, τ_w) orientation is identical to that when the plane is in the $(\sigma_w, \tau_w + \pi)$ orientation (see Eq. 1).

Figure 5a shows the estimated slant values from the orthographic method over the range of world slants using the MAP decision rule. The circles show each of the estimated slant values over the world slants and tilts range. The unbroken line represents the mean estimated slant. These results are similar to those in Fig. 5 from Witkin (1981); the model performs best at intermediate and large world slants but tends to over-estimate slant particularly at low world slant values.

Figure 5b also shows a polar plot of the world (open symbols) and estimated (filled symbols) slant and tilt pairs using the MAP decision rule. The increasing size of the symbols reflects the increasing slant conditions and allows easier differentiation between the estimates. Note that for the larger values of world slant the estimated slants are quite accurate (they lie on a circle with the same radius as the world slant in question). However, often the estimated tilt is around 180° away from the world tilt.

Figure 5c, d shows the corresponding results for the EXP decision rule. Note in Fig. 5c that the estimated slant values show marked underestimation of the corresponding world slants. This is due to the fact that the posterior has two

Fig. 5 Recovered slant and tilt estimates for the Witkin (1981) model using an orthographic projection method. **a** Slant estimates for MAP decision rule. **b** Mean slant and tilt estimates for MAP decision rule. **c** Slant estimates for EXP decision rule. **d** Mean slant and tilt estimates for EXP decision rule. In **a** and **c**, each circle corresponds to the slant recovered from one of 100 individual trials, the line connects the mean values over these 100 trials. In **b** and **d**, *open shapes* correspond to the simulated range of surface slants and tilts in the world. Closed symbols represent the recovered mean slant & tilt estimates. Note that the recovered tilt is often around 180° in error (which can be seen particularly clearly for the 0° tilt case)



peaks and consequently the spherical expectation in the slant dimension is between the peaks (and hence close to zero slant). The underestimation of slant can also be seen in the polar plot of Fig. 5d in which many of the estimated slants lie at the origin. Since the largest peak will sometimes be at the appropriate value of tilt and sometimes be 180° away from this value (due to random fluctuations in the projected orientations), Fig. 5d also shows the tendency for the method to misestimate tilt by 180° .

In this section, we have shown that the estimates obtained for the orthographic method of Witkin (1981) are relatively accurate for recovery of slant (provided slant is sufficiently large) but suffer from large misestimates of tilt. In the next section, we contrast the performance of the perspective method developed in this article.

3.2 Results—perspective projection method

Figure 6 shows contour plots for the set of posterior distributions averaged over the 100 trials, for each of the (σ_w, τ_w) conditions. Comparison with Fig. 4 indicates that there is a clear tendency for only one dominant peak using this method relative to the orthographic method. This is due to the fact that the perspective method developed here assesses the like-

lihood of observed orientations taking into account the line location on the plane. Consequently, only one of the peaks is favoured. Note also that, once again, as the slant is increased, the reliability of the estimate appears to improve (the spread of the peak decreases).

Figure 7a shows the estimated slant values from the perspective method over the range of world slants using the MAP decision rule. The circles show the estimated slant values over all world slants and tilts. The unbroken line represents the mean estimated slant. These results show a similar tendency to those in Fig. 5 from Witkin (1981) and similar to those seen in Fig. 5a of the present article; the model performs best at higher world slant values but tends to over-estimate slant at low world slant values. Note, however, that for the orthographic model the overestimate persists even in the highest slant condition tested. For the perspective model the overestimate vanishes at high slant values.

Figure 7b shows a polar plot of the world (open symbols) and estimated (filled symbols) slant and tilt pairs for the MAP decision rule. The distinction between these results and those from the orthographic method are now rather striking. The perspective method appears much better at estimating the world tilt.

Fig. 6 Contour plots derived from posterior distribution for the new perspective projection method model for a range of world slants and tilts (each subplot corresponds to a different world slant and tilt pair)... Note that now there is one dominant peak, thereby the ambiguity seen in Figs. 4 and 6 is reduced

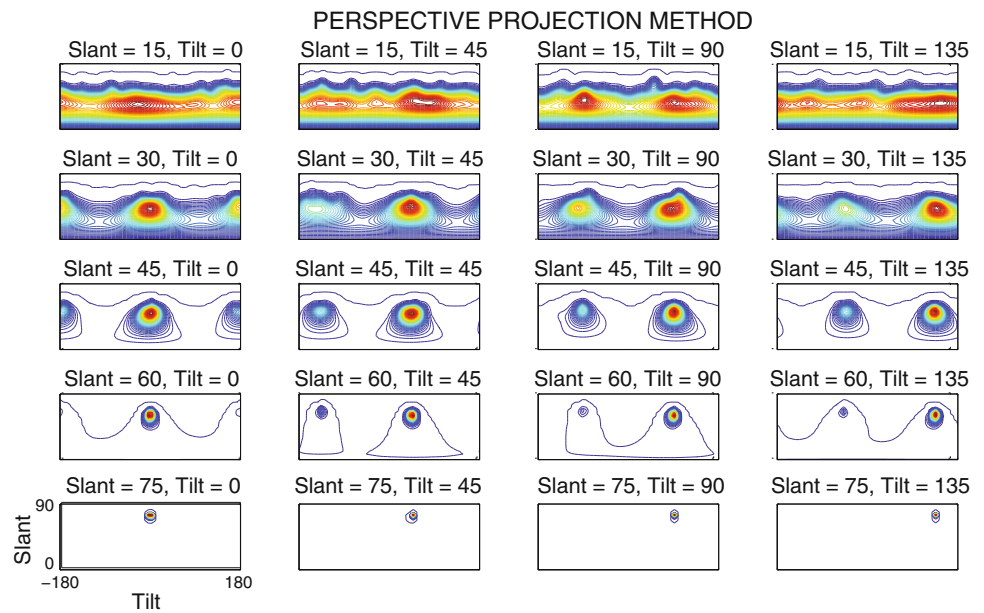


Fig. 7 Recovered slant and tilt estimates for the new perspective projection model. **a.** Slant estimates for MAP decision rule. **b.** Mean slant and tilt estimates for MAP decision rule. **c.** Slant estimates for EXP decision rule. **d.** Mean slant and tilt estimates for EXP decision rule. In **a** and **c**, each circle corresponds to the slant recovered from one of 100 individual trials, the line connects the mean values over these 100 trials. In **b** and **d**, open shapes correspond to the simulated range of surface slants and tilts in the world. Closed symbols represent the recovered mean slant & tilt estimates. Note that the recovered tilt is now much closer to the world tilt which corresponds to the down-weighting of the second peak in the posterior distribution in Fig. 6

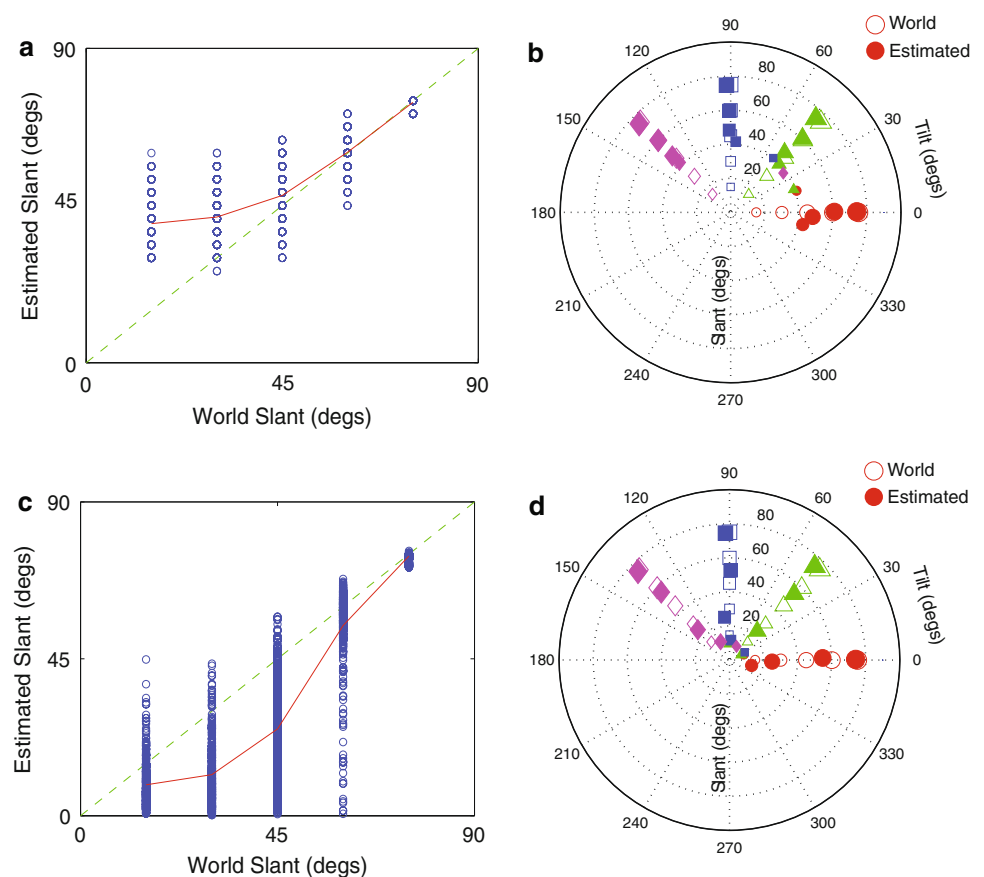
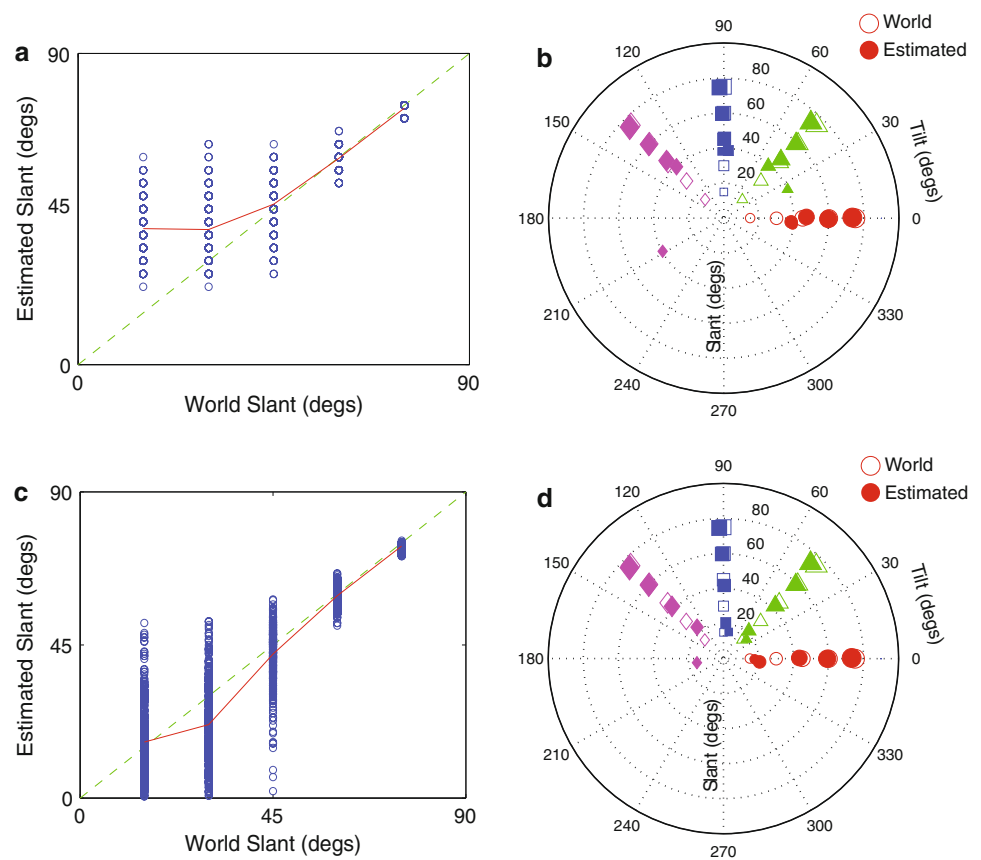


Figure 7c, d shows the corresponding results for the EXP decision rule. Note in Fig. 7c that the estimated slant values show much smaller under-estimates of slant relative to the results from the orthographic method (c.f. Fig. 5c). This is

due to the fact that the second peak in the posterior distribution is considerably smaller under the perspective method and consequently the spherical expectation is closer to the highest peak.

Fig. 8 Recovered slant and tilt estimates for the new perspective projection model when field of view is increased to 40°. **a.** Slant estimates for MAP decision rule. **b.** Mean slant and tilt estimates for MAP decision rule. **c.** Slant estimates for EXP decision rule. **d.** Mean slant and tilt estimates for EXP decision rule. In **a** and **c**, each circle corresponds to the slant recovered from one of 100 individual trials, the line connects the mean values over these 100 trials. In **b** and **d**, open shapes correspond to the simulated range of surface slants and tilts in the world. Closed symbols represent the recovered mean slant & tilt estimates. Note that errors in slant and tilt are now much smaller for the EXP decision rule (although they remain fairly similar to those seen in Fig. 7 for the MAP rule)



Relative to the orthographic simulation (Fig. 5d), Fig. 7d reveals much better approximations to the world slant and tilt values using the EXP decision rule under the perspective method. The tilts are accurately recovered and the slants show a similar pattern of underestimation to that shown in Fig. 7c.

In a further simulation, we show that the perspective method can produce very accurate results when the field of view is increased. We repeated the simulation shown in Figs. 6 and 7 but increased the field of view to 40°. To account for the increase in the texels now falling inside the aperture we reduced the number of texels on the surface to 2,500.

At larger fields of view the perspective information is greater and so we would expect the perspective method to perform particularly well in this condition. This expectation is confirmed to some extent in Fig. 8; when compared to the simulations conducted with small field of view and illustrated in Fig. 7 the MAP decision rule performs slightly better and the EXP decision rule performs much better. With these results in mind, it is worth noting that human observers have been shown to increasingly use perspective information at larger fields of view in the estimation of structure from motion (e.g., Eagle and Hogervorst 1999; Hogervorst and Eagle 2000).

3.3 Model comparison

We compared the orthographic and perspective projection methods and the EXP and MAP decision rules in a full factorial 2×2 design. We ran the simulations described in the previous sections (with 20° field of view) 10 times, recovering mean slant and tilt estimates over the 100 trials for each of the 20 world slant and tilt pairs and for each of the four models. For each of the 10 simulations we then calculated the RMS error in both estimated slant and tilt over the world slant and tilt pairs. Mean RMS slant and tilt error over the 10 simulations are shown in Fig. 9a and b, respectively. Note the error bars represent 95% confidence intervals and are smaller than the symbols used in Fig. 9a.

From Fig. 9 it is clear that the perspective models outperform the orthographic models. As noted in the previous section this is particularly clear for tilt estimation for which RMS errors are around 100° in the orthographic case.

Two 2-factor (projection model \times decision rule) ANOVAs, one for slant and one for tilt estimation were conducted. For slant estimation, there was a significant main effect of both projection model ($F(1, 9) = 53926.4$, $P < 0.0001$) and decision rule ($F(1, 9) = 24640.2$, $P < 0.0001$) and an interaction between these factors ($F(1, 9) = 22388.8$, $P <$

Fig. 9 Comparison of models using orthographic and perspective methods and the MAP and EXP decision rules. Mean RMS error in estimated **a.** slant and **b.** tilt over 10 replications. Error bars (which are smaller than the symbols in **a.**) represent 95% confidence intervals

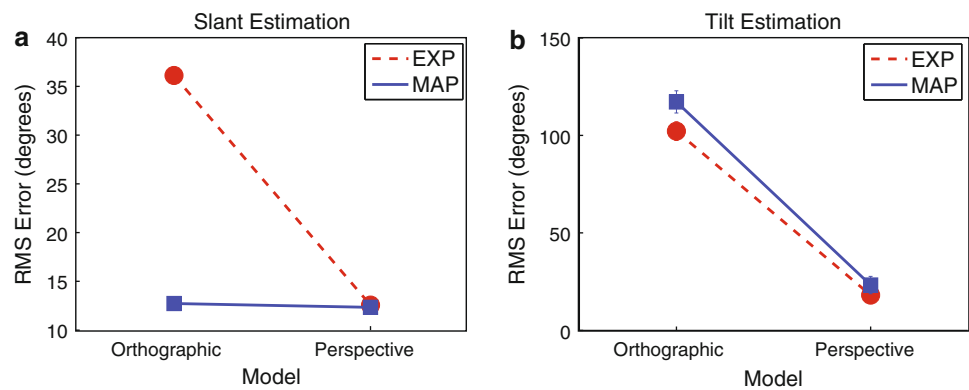


Table 1 Results of four repeated measures *t*-tests undertaken to assess differences between estimated slants in conditions sharing one common component (decision rule or projection method)

Comparison	<i>t</i>	df	<i>P</i> -value (2-tailed)
(ORTH, EXP) vs. (ORTH, MAP)	363.861	9	$P < 0.0001$
(PERSP, EXP) vs. (PERSP,MAP)	1.639	9	$P = 0.1356$
(ORTH, EXP) vs. (PERSP, EXP)	233.179	9	$P < 0.0001$
(ORTH, MAP) vs. (PERSP, MAP)	4.644	9	$P = 0.0012$

Table 2 Results of four repeated measures *t*-tests undertaken to assess differences between estimated tilts in conditions sharing one common component (decision rule or projection method)

Comparison	<i>t</i>	df	<i>P</i> -value (2-tailed)
(ORTH, EXP) vs. (ORTH, MAP)	−8.289	9	$P < 0.0001$
(PERSP, EXP) vs. (PERSP,MAP)	−2.252	9	$P = 0.0508$
(ORTH, EXP) vs. (PERSP, EXP)	27.487	9	$P < 0.0001$
(ORTH, MAP) vs. (PERSP, MAP)	31.777	9	$P < 0.0001$

0.0001). Repeated measures *t* tests were also conducted to assess the differences between the four conditions containing one common factor level. The results are presented in Table 1. Note that for both the MAP and EXP decision rules, the perspective model has significantly smaller error (although this improvement is not materially important in the case of the MAP decision rule). Note also that there is no difference in performance in the perspective condition between the MAP or EXP decision rules.

For tilt estimation there was a significant main effect of both projection model ($F(1, 9) = 1054.4$, $P < 0.0001$) and decision rule ($F(1, 9) = 37.124$, $P < 0.0002$) and an interaction between these factors ($F(1, 9) = 15.7$, $P < 0.005$). Similar to the analysis for slant, four repeated measures *t* tests were also conducted and the results are presented in Table 2. The results clearly indicate that the perspective model is better than the orthographic model and that this improvement is seen irrespective of whether the MAP or EXP decision rule is used.

Figure 10 shows the distributions of absolute errors in estimation of slant and tilt for each of the 4 models discussed.

Error frequencies were calculated over the 20 world slant and tilt pairs used in previous simulations and over 100 trials (i.e., 2,000 errors in total). The distributions of errors in slant estimation are shown in Fig. 10a. These are similar across models except for the model using the orthographic method and the EXP decision rule (bottom left panel) which tends to produce larger errors in recovered slant (this is expected from Fig. 5c). The distributions of errors in tilt estimation are shown in Fig. 10b. Note that the two models using the orthographic decision rule have peaks in their error distributions at 0° and 180° due to the inherent ambiguity in tilt recovery under this scheme. The tilt error distributions for the two perspective projection models also have a peak at zero but, in contrast to the orthographic case, the likelihood of obtaining an error in tilt estimation of 180° is greatly reduced.

Taken together, the data presented in Figs. 9 and 10 clearly indicate that the perspective projection method outlined in this article performs better than the orthographic method.

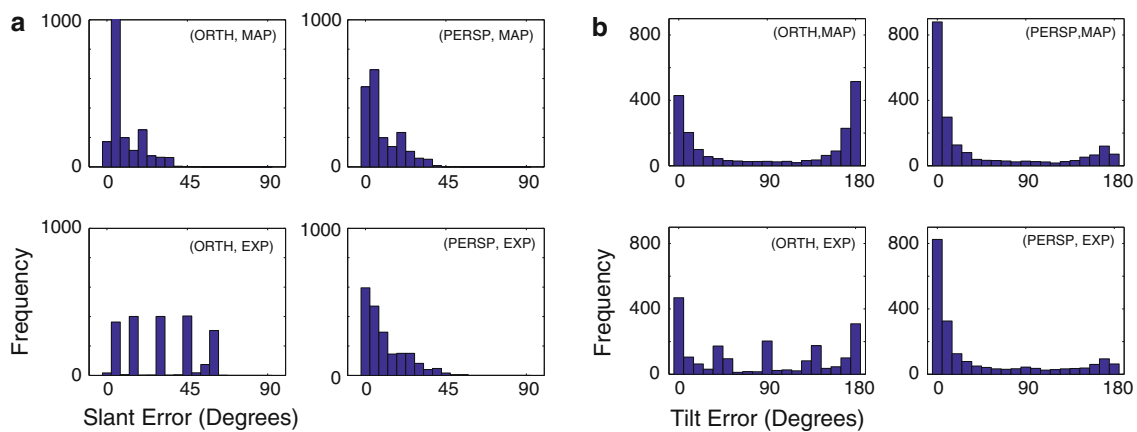


Fig. 10 Comparison of absolute errors in slant and tilt obtained using the orthographic and perspective methods and the MAP and EXP decision rules. **a.** Histogram of absolute slant errors. **b.** Histogram of absolute tilt errors

4 Discussion

4.1 Summary

We have derived a new model for the recovery of planar surface slant and tilt under perspective projection of oriented line segments. The model is a generalisation of the orthographic method presented in Witkin (1981) and relies upon assumptions of isotropy and homogeneity of texture elements on the plane. The new perspective model was compared to the Witkin (1981) model in a slant and tilt estimation task for a range of simulated surface pose parameters. The new perspective method outperformed the Witkin model and in particular it did not suffer from the ambiguity in surface tilt due to the orthographic projection process. An expected value decision rule rather than the MAP solution was also investigated. When the perspective projection method developed in this article was used, the model incorporating the expected value decision rule was found to have some similarities with human slant estimation in a human behavioural study (see supplementary materials).

4.2 Why are there errors?

Although the methods implemented produced reasonable estimates of slant and tilt (using either decision rule), there are still large and systematic deviations from the true value of surface orientation. It should, however, be noted that the performance is rather good bearing in mind that we are only considering the orientation of markings on the surface. There are many other sources of texture information available in the scene, as noted in the introduction (e.g., the density, size etc.). Adding further independent likelihood functions for these information sources would lead to smaller errors in estimated surface orientation.

4.3 Orthographic vs. perspective method

The new model based on perspective projection was shown to produce a more accurate approximation of surface pose than that of Witkin (1981). The results shown in Figs. 5a, b and 7a, b for the orthographic and perspective methods, respectively, and using the MAP decision rule, indicate that the perspective method is superior in terms of recovery of pairs of world slant and tilt values. In particular the perspective method is greatly superior in terms of estimation of tilt since the orthographic method regularly recovers tilt with a 180° error (Fig. 10).

In addition, the perspective method is also shown to outperform the orthographic method when a spherical expectation decision rule was implemented (Figs. 5c, d; 7c, d). Using this decision rule, the orthographic method greatly underestimates slant (in fact recovered slant is close to zero for values of world slant less than around 45° since the expectation lies approximately between the two peaks). Furthermore, the recovered tilt is still subject to systematic errors of around 180° caused by the inherent ambiguity present in the assumption of orthographic projection (Fig. 10b). In contrast, for the perspective method the spherical expectation decision rule produces a less pronounced underestimate of slant and accurate estimates of tilt. These findings are summarised in Figs. 9 and 10 which compare the four models implemented directly.

The results of the experiment undertaken in the supplementary materials indicate that human observers appear to be sensitive to the information arising from the perspective projection of orientated line segments. In particular, we found evidence for a significant increase in perceived slant when image orientations were defined by perspective orientation relative to orthographic projection (see supplemental materials). Consequently, we suggest that models of human surface pose recovery should incorporate a perspective projection process.

4.4 MAP vs. EXP decision rule

The MAP decision rule has the advantageous property that it minimises the expected error of the estimate. In this sense it is the optimal decision rule for our estimator to implement. However, it has been suggested that when human observers are faced with similar estimation problems and there is no explicit cost involved in making an incorrect decision then it is more appropriate to model the response by sampling from the posterior distribution (Mamassian and Landy 1998; Mamassian et al. 2002). In this case the spherical expectation should reflect the mean of the observer's responses.

Furthermore, it is possible to process slant and tilt independently to recover the MAP solution (i.e., first find the tilt at which the posterior is maximised and then examine all values of slant for this tilt value to find the MAP solution). In contrast the EXP solution can not be recovered in this way since it requires non-independent processing of slant and tilt. Consequently, unlike the EXP solution, use of a MAP solution is not consistent with the findings of Seyama et al. (2000). In this study, the authors tested whether adapting to slanted surfaces depended on the tilt difference between the adapting and test surfaces. They found that the slant aftereffect did in fact depend on tilt difference, suggesting non-independent processing of slant and tilt.

Commonly, human observers tend to underestimate surface slant when only texture information is present (e.g., see Gibson 1950; Braunstein 1968). As shown in Fig. 7 (and see Fig. S2 in supplemental materials), results obtained using the perspective method with a spherical expectation decision rule exhibit the moderate underestimation of slant associated with human observers. As a consequence this scheme may be able to explain some of this systematic underestimation.

4.5 Invalid assumptions for human observers?

The key assumption at the heart of both the orthographic and perspective method maximum likelihood estimators is that surface texture orientations are isotropic. Indeed there is some evidence that this is the case for human observers (Rosenholtz and Malik 1997; Knill 2003; Warren and Mamassian 2003), however, this assumption is clearly violated when, for example, one looks at a field of grass. Clearly, in such a situation the predominant orientation is vertical. This assumption, then may seem inappropriate for a model of human slant recovery. Knill (2003), however, has proposed and provided evidence for a framework in which prior assumptions such as isotropy and homogeneity are 'switched off' by human observers when there is sufficient information in the image to suggest that the assumption is invalid. This

results in a more general (less constrained) analysis of the world. Consequently, the fact that we can find examples of surfaces which violate key assumptions of the model presented here does not necessarily rule out its potential contribution to the description of human slant recovery.

References

- Blake A, Marinos C (1990) Shape from texture—estimation, isotropy and moments. *Artif Intell* 45:323–380
- Blake A, Bühlhoff HH, Sheinberg DL (1993) Shape from texture—ideal observers and human psychophysics. *Vis Res* 33:1723–1737
- Brady M, Yuille A (1984) An extremum principle for shape from contour. *IEEE Trans Pattern Anal Mach Intell* 6:288–301
- Braunstein ML (1968) Motion and texture as sources of slant information. *J Exp Psychol* 78:247–253
- Buckley D, Frisby JP, Blake A (1995) Does the human visual system implement an ideal observer theory of slant from texture? *Vis Res* 36:1163–1176
- Eagle RA, Hogervorst MA (1999) The role of perspective information in the recovery of 3D structure-from-motion. *Vis Res* 39:1713–1722
- Gårding J (1993a) Direct estimation of shape from texture. *IEEE Trans Pattern Anal Mach Intell* 15:1202–1208
- Gårding J (1993b) Shape from texture and contour by weak isotropy. *J Artif Intell* 64:243–297
- Gibson JJ (1950) *The perception of the visual world*. Houghton Mifflin, Boston
- Gruber HE, Clark WC (1956) Perception of slanted surfaces. *Percept Mot Ski* 6:97–106
- Hogervorst MA, Eagle RA (2000) The role of perspective effects and accelerations in perceived three-dimensional structure-from-motion. *J Exp Psychol Hum Percept Perform* 26:934–955
- Horn BKP, Brooks MJ (1989) *Shape from shading*. MIT Press, Cambridge, Mass
- Kanatani K (1984) Detection of surface orientation and motion from texture by a stereological technique. *Artif Intell* 23:213–237
- Knill DC (1998) Surface orientation from texture: ideal observers, generic observers and the information content of texture cues. *Vis Res* 38:1655–1682
- Knill DC (2003) Mixture models and the probabilistic structure of depth cues. *Vis Res* 43:831–854
- Koenderink JJ, van Doorn AJ, Kappers AML (1992) Surface perception in pictures. *Percept Psychophys* 52:487–496
- Li A, Zaidi Q (2000) Perception of three-dimensional shape from texture is based on patterns of oriented energy. *Vis Res* 40:217–242
- Malik J, Rosenholtz R (1997) Computing local surface orientation and shape from texture for curved surfaces. *Int J Comput Vis* 23:149–168
- Mamassian P, Landy MS (1998) Observer biases in the 3D interpretation of line drawings. *Vis Res* 38:2817–2832
- Mamassian P, Landy MS, Maloney LT (2002) Bayesian modelling of visual perception. In: Rao R, Olshausen B, Lewicki M (Eds.) *Probabilistic models of the brain: perception and neural function*. MIT Press, Cambridge, MA, pp. 13–36
- Mardia KV, Jupp PE (1999) *Directional statistics*, 2nd ed. Wiley, Chichester
- Mingolla E, Todd JT (1986) Perception of solid shape from shading. *Biol Cybern* 53:137–151
- Rosenholtz R, Malik J (1997) Surface orientation from texture: isotropy or homogeneity (or both)? *Vis Res* 37:2283–2293
- Seyama J, Takeuchi T, Sato T (2000) Tilt dependency of slant aftereffect. *Vis Res* 40:349–357

- Stone JV (1993) Shape from local and global analysis of texture. *Phil Trans R Soc Lond B* 339:53–65
- Velisavljevic L, Elder JH (2006) Texture properties affecting the accuracy of surface attitude judgements. *Vis Res* 46:2166–2191
- Warren PA, Mamassian P (2003) The dependence of slant perception on texture orientation statistics. *J Vis*, 3, abstract 847
- Witkin AP (1981) Recovering surface shape and orientation from texture. *Artif Intell* 17:17–45