Bayesian Modelling of Perception of Structure from Motion

Francis Colas[1], Pierre Bessière[2], Jacques Droulez[3], and Mark Wexler[3]

[1] INRIA Rhônes-Alpes - E-Motion
[2] CNRS - Grenoble Université
[3] Collège de France - LPPA

We use multiple sensory modalities to perceive our environment. One of these is optic flow, the displacement and deformation of the image on the retina. It is generally caused by a relative motion between an observer and the objects in the visual scene. As optic flow depends largely on three-dimensional (3D) shapes and motions, it can be used to extract structure from motion (the SFM problem). Motion parallax and the kinetic depth effect are special cases of this phenomenon, noticed by Von Helmholtz (1867), and experimentally quantified by Wallach and O'Connell (1953).

Extraction of shape from motion is a difficult issue for two main reasons: it is an ill-posed and inverse problem. Many different combinations of shape and motion can lead to the same optic flow, so reconstructing a shape from the optic flow cannot lead to a unique result. In this aspect, SFM is an ill-posed problem. Furthermore, geometry and optics can lead to the expression of optic flow given 3D structure and motion. However, SFM is interested in the opposite; that is, SFM is an inverse problem.

Previous approaches to the SFM problem rely mostly on the *rigidity assumption*, the hypothesis that optic flow is caused by the motion of a rigid object. Under this assumption, the number of degrees of freedom of the motion is greatly reduced, and it has been shown that little optic flow information is required to recover the motion and the structure of the object (Ullman, 1979, Mayhew and Longuet-Higgins, 1982). This assumption is supported by human performance in some psychophysical experiments Wallach and O'Connell (1953), Koenderik (1986). Some more recent models, relying on local velocity information rather than on the full optic flow field, are consistent with human performance (Todd and Bressan, 1990, Todd and Norman, 1991).

Extensive studies have investigated the perception by an immobile observer of optic flow caused by a mobile object. However, SFM can also occur with a static object for an observer in motion (Rogers and Graham, 1979). Until recently, it was believed that for the same optic flow, perception is the same with object motion or with subject motion (Wallach et al., 1974, Rogers and Graham, 1979). However, it has since been shown that the observer's movements influence the perceived 3D shape and motion (Rogers and Rogers, 1992, Dijkstra et al., 1995, Wexler et al., 2001b).

P. Bessière et al. (Eds.): Prob. Reason. & Deci. Mak., STAR 46, pp. 301–328, 2008.
springerlink.com

Therefore, the *stationarity assumption* has been introduced to account for the way that self-motion changes perception (Wexler et al., 2001b,a, Wexler, 2003). The stationarity assumption states that the visual system prefers the solution having minimal motion in an observer-independent, allocentric reference frame. Therefore, the stationarity assumption is the minimization of *absolute* motion, whereas the rigidity assumption can be seen as the minimization of *relative* motion between points of an object. Neither of these assumptions can explain human performance in the SFM task, and until now, no coherent model has been proposed to integrate these two assumptions.

In this chapter, we propose a generic Bayesian model that integrates the stationarity and rigidity assumptions for the perception of 3D planar surfaces from optic flow. The model accounts for the SFM performance of moving and stationary observers, as well as a number of other results reported in the literature. We begin with a generic model of SFM, then we give its instantiation for the perception of a plane. We finally compare the results of this model with six experiments found in the literature.

1 Model

1.1 Generic Model

The generic Bayesian model that we propose is a model of what an observer can deduce from the limited information received from optic flow. To cope with this limited information, this model uses probabilities to represent and handle the uncertainties that it faces. We construct our model according to the hypotheses evoked above. The first two are the stationarity (H1) and rigidity (H2) assumptions. Our model also assumes that the structure of the object is independent of both its motion and the motion of the observer (H3). We follow the Bayesian programming framework to specify this model (Lebeltel et al., 2004).

From relevant information to variables

For the general case of perception of an object by optic flow, we propose a model that takes into account: (i) the observed optic flow (denoted $\mathbf{\Phi}$); (ii) the 3D structure of the object (denoted Θ); (iii) the motion of the object (denoted X) in the observer's reference frame; (iv) the motion of the observer in the allocentric reference frame (denoted $\mathbf{M}$); and (v) the context of observation (denoted Λ).

From dependencies to decomposition

At the core of a Bayesian model lies the joint probability distribution over all its variables. This joint distribution is the expression of the hypotheses of a model. The structural part in the specification of the joint distribution summarizes the dependencies and independencies between the variables. This structure is called *decomposition*.

Hypothesis H1 is the rigidity assumption, which states that the observed optic flow is most likely that of a rigid object. As a consequence, the optic flow depends on the relative motion, the structure of the object, and the conditions of observation, but is independent of self-motion. This corresponds to the following mathematical simplification:

$$P(\mathbf{\Phi} \mid \Theta \wedge \mathbf{M} \wedge X \wedge \Lambda) = P(\mathbf{\Phi} \mid \Theta \wedge X \wedge \Lambda). \tag{1}$$

The stationarity assumption (H2) states that the object motion is most likely to be small in the allocentric reference frame. Therefore, the relative motion depends on self-motion. We use Bayes' rule to write:

$$P(\mathbf{M} \wedge X) = P(\mathbf{M})\, P(X \mid \mathbf{M}). \tag{2}$$

Hypothesis H3 states that the structure of the object is independent of both the relative motion of the object and self-motion. This translates as a product of independent factors in the decomposition:

$$P(\Theta \wedge \mathbf{M} \wedge X) = P(\Theta)\, P(\mathbf{M} \wedge X). \tag{3}$$

We also make the assumption that the conditions of observations are independent of the position of the plane as well as the motion of the object and self-motion. This is expressed as:

$$P(\Theta \wedge \mathbf{M} \wedge X \wedge \Lambda) = P(\Theta \wedge \mathbf{M} \wedge X)\, P(\Lambda). \tag{4}$$

Finally, using Bayes' rule, we can write:

$$P(\Theta \wedge \mathbf{M} \wedge X \wedge \Lambda \wedge \mathbf{\Phi}) = P(\Theta \wedge \mathbf{M} \wedge X \wedge \Lambda)\, P(\mathbf{\Phi} \mid \Theta \wedge \mathbf{M} \wedge X \wedge \Lambda). \tag{5}$$

Combining equations 5, 1, 3, 4, and 2, we obtain the decomposition shown in equation 6.

$$\begin{aligned} P(\Theta \wedge \mathbf{M} \wedge X \wedge \Lambda \wedge \mathbf{\Phi}) = {} & P(\Theta)\, P(\mathbf{M})\, P(\Lambda) \\ & \times P(X \mid \mathbf{M}) \\ & \times P(\mathbf{\Phi} \mid \Theta \wedge X \wedge \Lambda). \end{aligned} \tag{6}$$

This decomposition is the structural expression of our assumptions.

Physical and physiological laws → distributions

To derive a usable expression for the joint distribution, we must specify each of the factors in the above decomposition. The first factor, $P(\Theta)$, is the prior on the structure of the object. It represents what our model expects before any observation. It can be an uninformative prior or it can reflect some bias in the perception. The exact parametrical form depends on the actual experiment. In the same way, the second factor $P(\mathbf{M})$ represents the expectation by an observer of her or his own motion. If we assume that the model has an exact

knowledge of its self-motion (as will be the case later in this chapter), this probability distribution is simplified in the final inference and thus can be left unspecified. The same holds for the probability distribution over the conditions of observation $P(\Lambda)$. The inference will simplify this factor for any question in which the conditions of observations are exactly known.

The fourth factor, $P(X \mid \mathbf{M})$, specifies the relative motion expected from a given self-motion. According to stationarity, the object is more likely to undergo a smaller absolute motion. Therefore, the most probable relative motion should be defined as the opposite of self-motion. The actual parametrical form varies once again with the experiment, but a general expression could be proportional to the exponential of the opposite of kinetic energy (Gibbs distribution). In some cases (in an appropriate parameter space), this means a Gaussian distribution.

Finally, the last factor in decomposition 6 is the distribution on optic flow, given the structure of the object, the relative motion between the object and the observer, and the conditions of observation: $P(\mathbf{\Phi} \mid \Theta \wedge X)$. Following the rigidity assumption, this distribution states that the most probable optic flow is the theoretical flow of the object in this particular configuration, given this particular motion.

Formalized questions

A probabilistic question is a distribution over some variables of the model, possibly given the knowledge of the values of other variables. With a completely specified joint distribution, the answers to such questions can be mechanically inferred with both Bayes' and marginalization rules.

This model is designed for the perception of structure from optic flow. The first question that we can ask is the probability of the object structure or shape given the optic flow, the self-motion, and the conditions of observation, written as $P(\Theta \mid \boldsymbol{\varphi} \wedge \mathbf{m} \wedge \lambda)$.[1] This question is answered by the following expression, given by Bayesian inference.

$$P(\Theta \mid \boldsymbol{\varphi} \wedge \mathbf{m} \wedge \lambda) = \frac{P(\Theta) \sum_{x \in X} P(x \mid \mathbf{m}) P(\boldsymbol{\varphi} \mid \Theta \wedge x \wedge \lambda)}{\sum_{\theta \in \Theta} P(\theta) \sum_{x \in X} P(x \mid \mathbf{m}) P(\boldsymbol{\varphi} \mid \Theta \wedge x \wedge \lambda)} . \tag{7}$$

This is the computation that we have used to obtain the results shown later. Given our hypotheses, this distribution represents our knowledge about the structure of the object (including its relative position with respect to the observer) that one can infer from the observation of optic flow and self-motion.

Furthermore, with this model, one can ask other questions. For example, one may be interested in the estimation of self-motion from optic flow: $P(\mathbf{M} \mid \boldsymbol{\varphi})$. This question can be used to study vection, in which optic flow induces the sensation of self-motion. In the same manner, one can also study the direction of perceived self-motion, called heading. With a single Bayesian model, different issues can be investigated.

[1] We use an upper case letter for a variable and lower case for the instantiation of a variable with a particular value.

1.2 Perception of a Plane

We can adapt this model to account for given situations. In this chapter, we consider a set of experiments dealing with the perception of a planar object through its optic flow. The observation is monocular, and the participant is sometimes allowed to move along some degrees of freedom while maintaining fixation. The observation can take place in either a small or large field of view.

With this information, we instantiate the template model following the Bayesian programming framework.

Variables

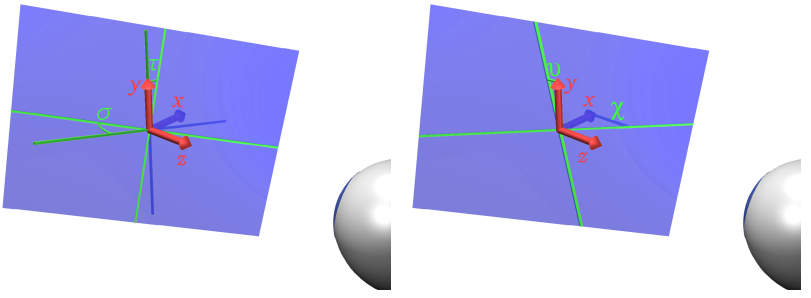
The first variable of the model is the optic flow. We must choose a representation space adapted to the specifics of the experiments. When observing a plane, the optic flow is uniquely determined by the derivatives of the image velocity with respect to the image coordinates, up to the second order (Longuet-Higgins, 1984). Therefore the chosen semantics and domain of $\mathbf{\Phi}$ are these derivatives. We can separate the optic flow into multiple components depending on the order of derivation: $\mathbf{\Phi} = \mathbf{\Phi}^0 \wedge \mathbf{\Phi}^1 \wedge \mathbf{\Phi}^2$.

As we consider a plane, the 3D structure of the object is restricted to its position. The participant maintains fixation on a given point of the plane. This way, only two parameters are required to specify the structure Θ. The extrinsic orientation of a plane in 3D space is often parameterized by two angles, slant and tilt (see Fig. 1(a)). Slant is the angle, in 3D space, between the plane's normal vector and the normal of the fronto-parallel plane. Tilt is the angle, in the fronto-parallel plane, of the projection of the plane's normal. For practical reasons in the computations, we use the depth gradients along both the transversal and vertical axes (respectively X and Υ), as depicted in Fig. 1(b).

The motion X of the object in the reference frame of the observer is defined by its rotation $\mathbf{\Omega}$ and translation $\mathbf{T}$ components along the transversal, vertical ang sagittal axes. Likewise, self-motion $\mathbf{M}$ is a set of translation and rotation velocities of the eye of the observer. Finally, the context of observation Λ reduces, for the experiments considered, to the size of the field of view, reduced to two extreme cases: small and large fields of view.

Decomposition

The instantiated model inherits the conditional independencies of the generic one. However, we can add some new simplifications in the joint distribution, depending on our knowledge of our experiments. In this case, the choice of parameterization of the optic flow allows us to make each component independent of the others, conditionally to knowledge of the relative position, relative motion and conditions of observation:



(a) (b)

Fig. 1. Illustration of the parameters of a plane. The object plane is shown in blue; the (x, y, z) reference frame is shown in red, and the observer (shown by its eye) is located at $(0, 0, d > 0)$. (a) Illustration of the tilt τ and slant σ of a plane. The slant is the angle between the object plane and the fronto-parallel plane. The tilt is the angle between the vertical axis and the intersection of the object plane and the fronto-parallel plane. (b) Illustration of parameters (X, Υ) of a plane. In green, we show the intersections of the object plane with both horizontal and sagittal planes. The respective slopes of these lines are the X and Υ parameters of the plane.

$$
\begin{aligned}
& P(\mathbf{\Phi} \mid \Theta \wedge X \wedge \Lambda) \\
= \; & P(\mathbf{\Phi}^0 \mid \Theta \wedge X \wedge \Lambda) \\
\times \; & P(\mathbf{\Phi}^1 \mid \Theta \wedge X \wedge \Lambda) \\
\times \; & P(\mathbf{\Phi}^2 \mid \Theta \wedge X \wedge \Lambda).
\end{aligned}
$$

Furthermore, we can state that only the perception of second-order optic flow depends on the size of the field of vision and that constant optic flow is generated only by the relative translation in the fronto-parallel plane. Finally our operative decomposition is:

$$
\begin{aligned}
P(\Theta \wedge \mathbf{M} \wedge X \wedge \Lambda \wedge \mathbf{\Phi}) = \; & P(\Theta)\, P(\mathbf{M})\, P(\Lambda) \\
& \times P(X \mid \mathbf{M}) \\
& \times P(\mathbf{\Phi}^0 \mid \mathbf{T}) \\
& \times P(\mathbf{\Phi}^1 \mid \Theta \wedge X) \\
& \times P(\mathbf{\Phi}^2 \mid \Theta \wedge X \wedge \Lambda).
\end{aligned} \tag{8}
$$

Distributions

As in the general case, we specify an exact parametrical form for each factor of our decomposition. The difference resides in our knowledge of our particular experiments.

We chose the prior on the position of the plane to be uninformative. That is, we state that each position of the plane is equally probable. In terms of probability distributions, this means that it is invariant by rotation around the

fixation point. For a given position $\theta = (\chi, \upsilon)$ of the plane, the probability density becomes:

$$P(\chi \wedge \upsilon) = \frac{1}{2\pi}\left(1 + \chi^2 + \upsilon^2\right)^{-\frac{3}{2}}.$$

As in the general model, the factors $P(\mathbf{M})$ and $P(\Lambda)$ can be left unspecified.

For $P(X \mid \mathbf{M})$, the expression of stationarity, we chose a Gibbs distribution with kinetic energy computed in the global reference frame. As X is expressed as the linear and angular velocities with respect to each axis, this means that $P(X \mid \mathbf{M})$ is a product of Gaussian distributions.

Finally, rigidity is expressed by the factors $P(\mathbf{\Phi}^0 \mid \mathbf{T})$, $P(\mathbf{\Phi}^1 \mid \Theta \wedge X)$, and $P(\mathbf{\Phi}^2 \mid \Theta \wedge X \wedge \Lambda)$. These distributions are chosen as Gaussian around the exact values of the optic flow computed by the theoretical equations (detailed in appendix A). The influence of the size of the field of view in the distribution of the second-order optic flow is in the covariance matrix (see the implementation for details).

Question

This is a model of the perception of a plane from optic flow. The question that we ask, to compare our model with the experimental results, is: what knowledge about the relative position of the plane can be deduced from the observation of optic flow, self-motion and the conditions of observation? As noted above, this becomes, in probabilistic terms:

$$P(\Theta \mid \boldsymbol{\varphi} \wedge \mathbf{m} \wedge \lambda).$$

With a more specific decomposition than the generic model, we can write the answer to this question with greater precision:

$$\begin{aligned} &P(\Theta \mid \boldsymbol{\varphi} \wedge \mathbf{m} \wedge \lambda) \\ &\propto P(\Theta) \sum_{\mathbf{t} \in \mathbf{T}} P(\boldsymbol{\varphi}^0 \mid \mathbf{t}) \sum_{\boldsymbol{\omega} \in \boldsymbol{\Omega}} P(x \mid \mathbf{m}) P(\boldsymbol{\varphi}^1 \mid \Theta\ X) P(\boldsymbol{\varphi}^2 \mid \Theta \wedge X \wedge \Lambda). \quad (9) \end{aligned}$$

Implementation

For the instantiated model, a number of parameters are necessary for the calculations.

With the notable exception of Gaussian distributions in some particular conditions (linearity for instance, which is not the case here; see the appendix), Bayesian inference is usually done with discretized or sampled variables. Table 1 gives the details of the ranges (minimum, maximum and number of samples in between) and dimensionality of each component of Θ (top row), of the relative rotation (middle row), and of the relative translation (bottom row). Other variables do not require discretization as their values are known for the inference.

Moreover, some of the distributions in our decomposition involve parameters; for instance, the Gaussians on relative motion and optic flow. Table 2 presents their respective covariance matrices. We notice that each one is diagonal and

Table 1. Domains of the variables

Variable	symbol	min	max	values by dimension	dimensions
Depth gradient	Θ	-4.125	4.125	33	2
Angular velocity	Ω	$-1.375\ rad.s^{-1}$	$1.375\ rad.s^{-1}$	11	3
Linear velocity	T	$-1.375\ m.s^{-1}$	$1.375\ m.s^{-1}$	11	3

Table 2. Covariance matrices for each factor of the joint distribution. From top to bottom: distribution over the relative translation, relative rotation, order-0 optic flow, order-1 optic flow, order-2 optic flow in a small field of vision and order-2 optic flow in a large field of vision. $Id_{n\times p}$ stands for the identity matrix of size n by p.

Distribution parameters
$\sigma_{\mathbf{T}} = 0.3 * Id_{3\times 3}$ in $m.s^{-1}$
$\sigma_{\mathbf{\Omega}} = 1.2 * Id_{3\times 3}$ in $rad.s^{-1}$
$\sigma_{\Phi^0} = 1.0 * Id_{2\times 2}$ in $m.s^{-1}$
$\sigma_{\Phi^1} = 0.025 * Id_{4\times 4}$ in s^{-1}
$\sigma_{\Phi^2 \mid \lambda=SF} = 5.0 * Id_{2\times 2}$ in $m^{-1}.s^{-1}$
$\sigma_{\Phi^2 \mid \lambda=LF} = 0.2 * Id_{2\times 2}$ in $m^{-1}.s^{-1}$

that the covariance on optic flow is greater in a small field of view for second-order optic flow than in a large field. Indeed, because second-order optic flow is quadratic in the distance from the fixation point, it is harder to perceive in a small field of vision than in a large field.

The results presented in the following section were all computed with this single set of parameters using the ProBT inference engine for the calculations (Lebeltel et al., 2004).

2 Results

In this section, we present six psychophysics experiments concerned with the monocular perception of a rotating planar patch with a neutral or non-informative texture. Therefore, the only cue for the plane's relative position was its motion via the optic flow. These experiments were previously reported by different teams and admit some variations around a common set-up, involving the motion of the observer's head or eyes, or the plane, or both, as well as the size of the displayed stimulus. This is in accordance with the choice of the variables of our model. For each experiment, we present the results of the Bayesian model and compare its performance with the participants'.

2.1 Protocol

All these experiments share the same basic conditions of experimentation. They were carried out by different teams, and the variations mostly concern the testing of the various assumptions. Here, we present the common features, and we will detail the differences when necessary.

The stimulus is an animation showing a 3D slanted plane in rotation. This animation is displayed on a monitor or on a screen. The plane is figured by a random-dot texture, considered to be uninformative concerning the slant and tilt of the plane (see Fig. 2). The size of the display depends on the condition of observation. It can be either a small field of vision (inside a cone of 5 to 10° of semi-angle) or a large field of vision (around 60°).

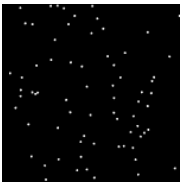


Fig. 2. Example of a frame of the animation presented to the participants. The display is composed of white dots of the object plane, projected on a black background. As the plane moves in 3D, the dots move according to their position on the object plane. The dots are chosen so that the distribution of their projection is invariant with respect to a 2D planar rotation.

The observation is monocular: the non-dominant eye of the participant is masked. The participants are also fitted with sensors so the experimenter can measure the position of their head or eye. Therefore, their motion can be used to check whether it is not too important (if they are asked not to move) and to change the projection of the dots of the object plane on the screen according to their changing point of view. The stimulus is displayed for less than 1 s.

The participants are asked to judge the slant or the tilt of the plane. To do so, after the presentation of the stimulus, they must align a probe (like a planar grid) using an input device (a joystick for example) with the mean position of the perceived plane.

2.2 Depth Reversal

Depth reversal is a well-known effect in 3D vision: many depth cues are ambiguous as to the sign of relative depth (cf. the Necker cube). In SFM, the simplest instance of this ambiguity is the observation of a rotating plane through a small opening. In this case, there is ambiguity regarding the tilt and direction of rotation, as illustrated in Fig. 3(b). However, it has been shown (Dijkstra et al.,

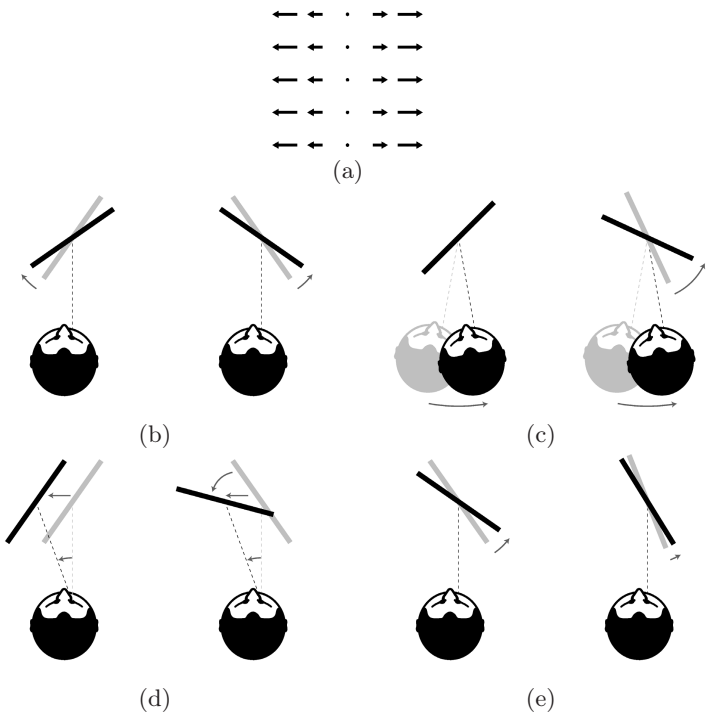


Fig. 3. Some ambiguities in first-order optic flow that have been used in the studies cited. (a) An example of an optic flow field that presents a number of ambiguities: all configurations shown in this figure lead to this flow. (b) These two configurations, which differ by simultaneous reversals of relative depth and 3D motion, both yield the optic flow shown in (a). This ambiguity is called depth reversal. (c) Depth reversals can also occur for moving observers. These two configurations have the same relative motion between object and observer as in (b) and therefore yield the same optic flow. However, one solution is stationary in an allocentric or observer-independent reference frame, while the other solution undergoes a rotation in this frame, twice as fast as the observer's motion. (d) The same ambiguity when the observer tracks a moving surface with the eyes. One solution undergoes a translation only, while the other undergoes the same translation and a rotation. (e) Ambiguity between slant and rotation speed: a larger slant coupled with a slower rotation may give the same optic flow as a lower slant together with a faster rotation.

1995) that this ambiguity does not occur with a large field of vision. We investigate this simple effect as a first example of our model.

The experiment that we use as a reference was described by Cornilleau-Pérès et al. (2002). In this particular experiment, the participants are asked not to move. The optic flow is induced by the plane, in rotation around an axis in the fronto-parallel plane.

Cornilleau-Pérès et al. (2002) report the results in terms of the rate of tilt reversals. A tilt reversal is defined to occur when the absolute error in the estimation of the tilt angle is greater than 90°. The reversal rate can be considered as a measure of the ambiguity, as illustrated in Fig. 3(b). The middle column of Table 3 presents the results of the experiment, and we observe that the reversal rate drops from close to its maximal value (50%) in a small field of vision to below 5% in a large field of vision.

Table 3. Influence of the size of field of vision on the reversal rate. Both the experiment (Cornilleau-Pérès et al., 2002) and the Bayesian model exhibit smaller reversal percepts in a large field of vision.

Condition	Experiment	Model
Small field	48.8%	44.6%
Large field	3.1%	3.3%

As detailed above, the Bayesian model computes the probability distribution over the orientation Θ of the plane, given the optic flow, the field of view and the observer's movement (example in Fig. 4). An ambiguity in the optic flow interpretation, such as the one illustrated in Fig. 3, results in a multimodal probability distribution. To compare the reversal rate reported by Cornilleau-Pérès et al. (2002) with the model output, we compute the sum of probabilities corresponding to tilt errors greater than 90° (see Table 3).

This result is accounted for by the rigidity assumption. In our model, this assumption is expressed by a probability distribution over the optic flow. The tilt ambiguity is a consequence of the invariance of the first-order components of optic flow ($\mathbf{\Phi}^1$) with respect to tilt reversal; therefore, only the second-order components can disambiguate the stimulus.

The reduction of reversal rate in a larger field is explained in the model by the increased influence of second-order optic flow components on the speed of a dot, which is quadratic in the size of the field of view. In contrast, the influence of first-order optic flow is only linear. Therefore, the relative influence of second-order versus first-order optic flow increases with the size of field. In our model, this is expressed by a reduced uncertainty of the second-order optic flow for large fields of view.

Qualitatively, in so far as this uncertainty is greater in small fields, the probability of depth reversal will always be higher in a small field than in a large field. Figure 5 shows the quantitative evolution of the reversal rate in the model as a function of this parameter.

2.3 Depth Reversal in Moving and Immobile Observers

Recently, self-motion has been shown to modify depth perception from optic flow. This can be seen most clearly in studies that find differences in SFM performance

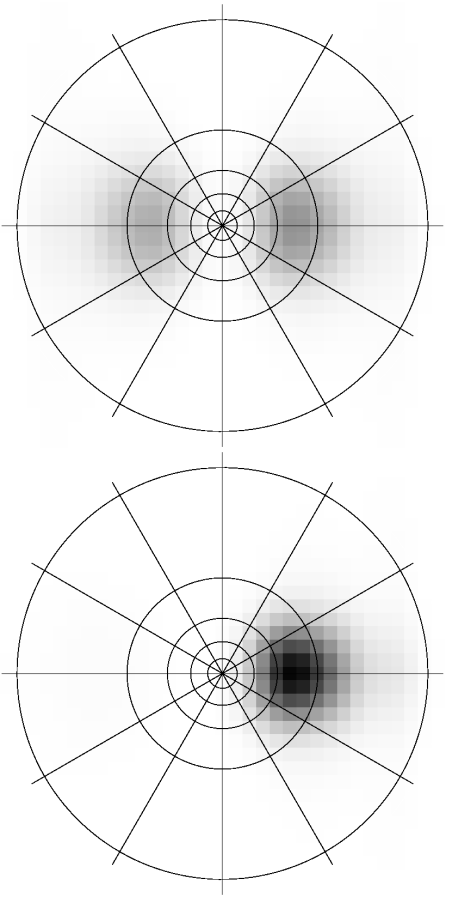


Fig. 4. Examples of probability distributions on the orientation of a plane. The polar angle is the tilt of the plane, the radius is the tangent of the slant angle, and the colour stands for the probability, with darker colours representing higher probabilities. The peaks represent the most likely percepts, with the integral of the probability around a peak corresponding to the probability of the associated percept. The top panel shows a result with a high rate of depth reversals, while the bottom result displays a low reversal rate.

between moving and immobile observers, while keeping optic flow the same in the two conditions. Thus, actively generated optic flow can lead to a different perception of 3D shape than the same optic flow viewed passively by an immobile observer.

One of the ways in which self-motion modifies SFM is by diminishing the ambiguity that leads to depth reversals (Rogers and Rogers, 1992, Dijkstra et al., 1995, Wexler et al., 2001a,b). An optic flow field such as the one shown in Fig. 3(a) leads, for an immobile observer, to total ambiguity between the solutions

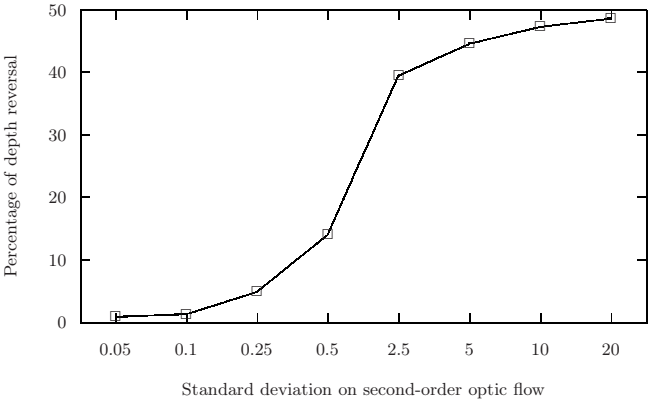


Fig. 5. Influence of the uncertainty of second-order optic flow on the reversal rate in the Bayesian model. A small field of vision leads to a greater uncertainty, and hence to more depth reversals.

shown in Fig. 3(b), and therefore a depth reversal rate of up to 50% for a small field of view. On the other hand, for a moving observer (Fig. 3(c)), the ambiguity is reduced in favour of the solution that is most stationary in an observer-independent reference frame (left solution in Fig. 3(c)).

The experimental data used as a reference are taken from Van Boxtel et al. (2003). The plane is observed in a small field of vision. There are two different conditions corresponding to Figs 3(b) and (c): either the participant is immobile and the plane is in rotation around the vertical axis (immobile condition), or the participant is moving his or her head transversally, and the plane is immobile in the allocentric reference frame (active condition).

The experimental results for both the active and immobile conditions are shown in Fig. 6. They reveal a bimodal distribution of tilt perception when the subject is immobile. There are two preferred responses: around 0°, corresponding to the simulated plane, and 180°, corresponding to the depth-reversed plane. In the active condition, the same optic flow is produced by the participant's displacement in front of an immobile plane. In this case, the depth-reversed plane is rarely reported, leading to a dominant peak in the distribution around 0°.

Figure 7 shows the results of our model in the same two situations. We notice the bimodality in the immobile condition similar to the experimental results as well as the decrease of depth reversals in the active condition. In the Bayesian model, the bimodality derives from the symmetry of first-order optic flow mentioned above. Furthermore, the difference between active and immobile conditions can be accounted for only by the conditional distribution on motion in an observer-independent reference frame. This distribution is the expression of the stationarity assumption in our model. In the immobile condition, the simulated and depth-reversed planes have the same speed, as depicted in Fig. 3(b); only the direction of motion changes. In the active condition, however,

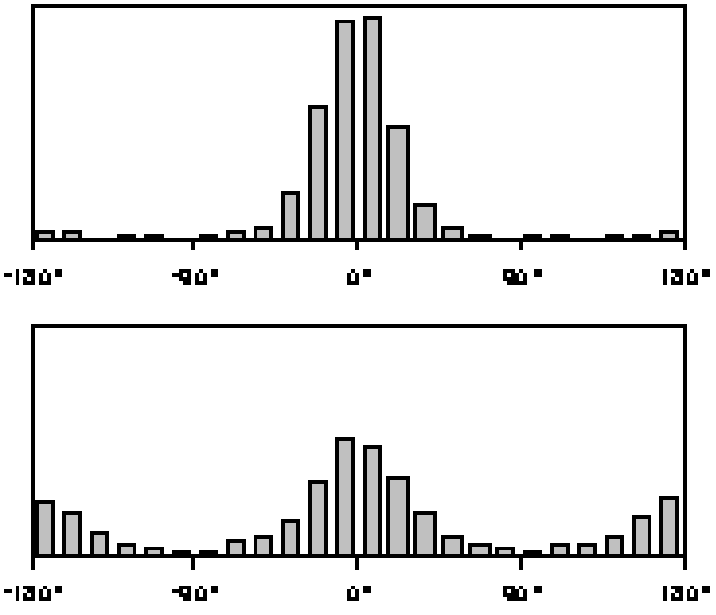


Fig. 6. Distributions of error on tilt angle for both active (top) and immobile (bottom) conditions, by Van Boxtel et al. (2003). The results show depth reversals in the immobile condition and its almost complete disappearance in the active condition.

the simulated plane is stationary in an observer-independent reference frame, whereas the depth-reversed plane has a high velocity (see Fig. 3(c)). Therefore, the stationarity assumption, as implemented in the model, ensures that the reversed plane is less probable, because it corresponds to a higher velocity in an observer-independent reference frame.

2.4 Ambiguity between Slant and Speed

The slant of a plane is difficult to extract from optic flow. Indeed, the rotation around an axis lying in the fronto-parallel plane is entangled with surface slant (the angle between the surface normal and the direction of gaze). Starting from a given slant and motion configuration, simultaneously increasing slant and decreasing motion leads to approximately the same optic flow.

The experimental data that we consider are taken from Domini and Caudek (1999). The experimental conditions involve a static monocular observer. The stimulus consists of a plane rotating along a fronto-parallel axis. The observer is asked to make a judgement about the slant of the plane. The planes can have two different slants and two different angular velocities. The relationship between the chosen slants is such that the tangent of the second slant is twice that of the first. The same holds for the velocity, where the second is twice that of the first (see Fig. 3(e)).

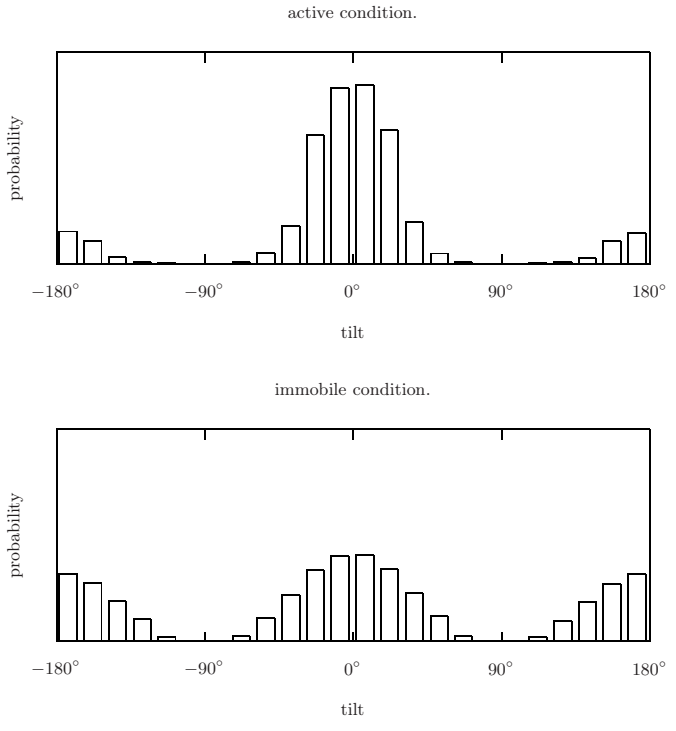


Fig. 7. Probability distributions of tilt errors in active and immobile conditions. As in the experimental results shown in Fig. 6, the ambiguity diminishes drastically in the active condition.

The experimental results, from Domini and Caudek (1999), are shown in Table 4. The columns on the left show the evolution of the perception of the tangent of the slant angle while changing the values of angular speed or the simulated slant. These data show that the slant of the plane is hardly recovered as an independent variable, arguing against a veridical (Euclidean, see the review by Domini and Caudek (2003)) analysis of optic flow by human observers. Moreover, the perceived slant for small simulated slant and high angular speed is very close to the one perceived in the case of large simulated slant at low speed. Finally, this experiment shows that increasing the simulated slant or increasing the angular speed yields the same increase in perceived slant (around 23% each time).

The right columns of Table 4 show the predictions of our model in the same experimental conditions. Our model shows the slant/speed ambiguity found in the experimental results. In particular, the perceived slant for small slant with high angular speed is very close to the perceived slant for large slant with low angular speed. These results also show an increase in slant perception while increasing either slant or speed. As in the experimental data, this increase is

roughly the same (50 to 60%) in both conditions, although greater than in the experimental data.

The perceived slant comes from a trade-off between our prior over the orientation (tilt and slant) of the plane and the distribution over the relative motion from the stationarity assumption (see the methods section for details).

Note that the values of perceived slant from the model are slightly smaller than those from the experimental data, especially for a small simulated slant. We have chosen to provide the results of our model with a unique set of parameters for all the experiments of this section. These parameters are therefore a trade-off between the best parameters fitting each experiment.

Table 4. Mean perceived tangent of slant as a function of simulated slant tangent and angular speed for both the experimental data (Domini and Caudek, 1999) and the Bayesian model. Note the growth of perceived slant with increasing angular speed and very similar perceived slant for large simulated slant/slow rotation and small simulated slant/fast rotation.

	Experiment		Model	
Angular speed	0.25	0.5	0.25	0.5
Small slant (1.5)	1.13	1.29	0.66	1.00
Large slant (3)	1.28	1.71	1.00	1.64

The slant/speed ambiguity results from ambiguities in first-order optic flow. In both situations (small slant, high speed compared to large slant, low speed) the optic flow is the same up to the first order, as shown in Fig. 3(e) and only the second-order optic flow could disambiguate the stimuli. These results confirm the low weighting of second-order components of optic flow in a small field of view. This low weighting arises from the uncertainty attached to the distribution over the second- order optic flow.

First-order optic flow depends on a parameter called *def*, the product of the tangent of the slant and angular speed (Domini and Caudek, 2003).[2] Therefore, slant and speed cannot be recovered individually from first-order optic flow. Domini and Caudek (2003) propose a maximum-likelihood model to account for their psychophysical results. With a small size of field, in the absence of self-motion and translation, and disregarding second-order optic flow, the likelihood of our Bayesian model reduces to the Gaussian $P(\mathbf{\Phi}^1 \mid \mathbf{\Omega} \wedge \Theta)$. The norm of the first-order optic flow in this case is $\sqrt{\omega_X^2 + \omega_Y^2}\sqrt{\chi^2 + \upsilon^2} = |\mathbf{\Omega}| \tan\sigma$. Their model is thus a special case of our Bayesian model.

2.5 Ambiguity of Translation in Depth

Another symmetry or ambiguity of first-order optic flow is shown in Fig. 8. A rotation in depth generates the same (up to first-order) optic flow as a transla-

[2] Projected on vertical and transversal axes, *def* is $\chi\omega_y$, $\upsilon\omega_y$, $\chi\omega_x$, $\upsilon\omega_x$ in the equations shown in the appendix.

tion in depth together with a different rotation in depth, around an axis that differs by 90° from the original rotation. It has been found (Wexler et al., 2001a, Wexler, 2003) that the two solutions are perceived with different frequencies, depending on the observer's movement and the origin of the depth translation—whether the observer moves towards the surface, or the surface moves towards the observer (see Fig. 8). These results can be summarized by stating that there is a strong bias towards perceiving the solution that minimizes motion in an observer-independent reference frame. Thus, these results provide further support for the stationarity assumption. However, the observer's percepts are also, by and large, in agreement with the rigidity assumption. Therefore, they provide a useful testing ground for our model, which incorporates both the stationarity and rigidity assumptions.

In the psychophysical studies, two conditions were tested: in the active condition, the observer moves his or her head in depth; in the immobile condition, the observer remains still but receives the same optic flow as in a previous active trial (Wexler et al., 2001a, Wexler, 2003).[3] In the active condition, the optic flow is generated by a plane rotating in depth at a fixed distance from the observer (the plane's centre therefore undergoes depth translation as well). Therefore, in the active condition 8(d), the rigidity assumption favours the simulated plane, while the stationarity assumption favours the alternative solution.[4] In the immobile condition, on the other hand, both the rigidity and stationarity assumptions favour the simulated plane.

The experimental results are presented in Fig. 8(c″), (d″), and in Table 5 as the fraction of trials in which the observers perceive the alternative, non-rigid plane. Recall that optic flow is the same in the active and immobile conditions; only the observer's motion differs. Providing that only first-order optic flow components are available, the rigidity assumption alone would predict equally low rates for the alternative solutions in the two conditions, whereas stationarity alone would result in a rate close to 100% in the active condition and a low rate in the immobile condition. Second-order optic flow components, if available, would decrease the rate for the alternative non-rigid solution.

As explained above, the discrepancy between the actual values of the experimental results and the model are because of the unique parameter set used for all six experiments. More precisely, different groups of participants already exhibit differences in their results. See, for instance, Fig. 8(c″) and the bottom-left histogram in Fig. 10. Both correspond to the same conditions, but the results are numerically different. Priors in our model can be adjusted to fit some results better at the expense of other experiments.

[3] Other conditions, involving conflict between the observer's motor command and self-motion, were also tested (Wexler, 2003) and were found to lead to different response distributions.

[4] The reason why the rigidity assumption favours the simulated plane rather than the alternative solution is that the symmetry of Fig. 8 only holds for first-order optic flow. Second-order terms break the symmetry and lead to non-rigidity of the alternative solution.

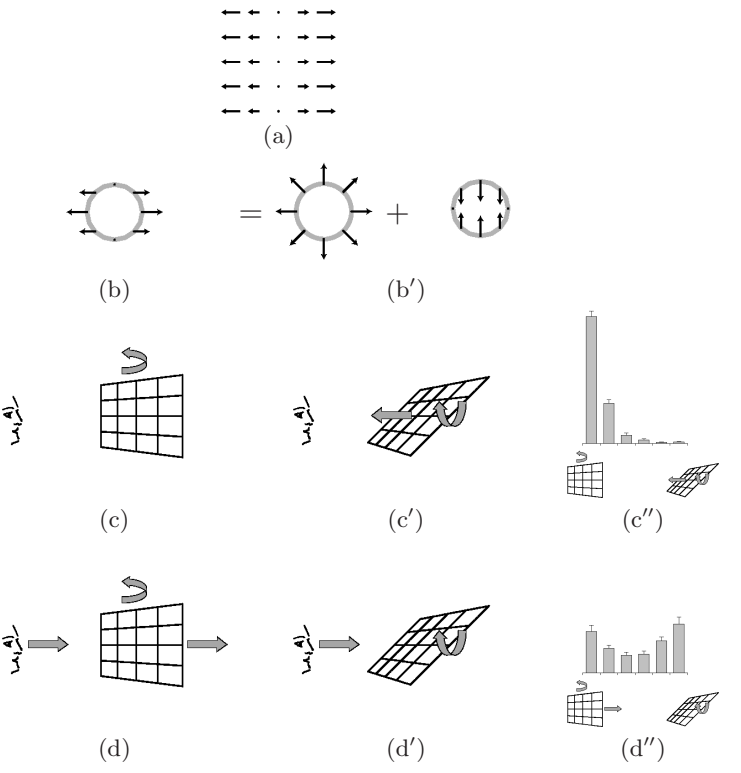


Fig. 8. Illustration of the effect of head motion on the perception of 3D structure (Wexler et al., 2001a, Wexler, 2003). (a) An ambiguous 2D optic flow field that can have different 3D interpretations, discovered by J. Droulez (cf. Fig. 3(a)). The arrows represent motions of projections of points in 3D space on the retina. It is fairly easy to see that the 3D configuration shown in (c) will generate this flow. However, the configuration shown in (c′) can also generate the flow in (a), and the reason for this is shown in (b) and (b′): if the amplitudes of the translation and rotation in (c′) are adjusted correctly, the rotation can exactly cancel the expansion flow from the depth translation in one of two dimensions. The planes in (c) and (c′) have the same slant and angular speed but different tilts, and they rotate about different axes. (d), (d′) Because optic flow depends only on the *relative* motion between object and observer, the same ambiguity holds for an observer moving forward and experiencing the optic flow in (a). If the observer's speed is equal and opposite to the translation in (c′), the stationarity of the solutions is reversed with respect to (c) and (c′): it is now the centre of (d′) that is stationary in space, while (d) translates at the same speed as the observer. (c″), (d″) Data from Wexler (2003), showing the frequencies of the perceived solutions for stationary (c″) and moving (d″) observers, with the bars on the left corresponding to solutions (c) and (d), and the bars on the right to solutions (c′) and (d′). Although the optic flow is the same in the two cases, perceptions of 3D structure are very different, showing the effect of the observer's action.

Table 5. Rate of alternative, non-rigid responses for the ambiguous depth-translation stimulus. Experimental results from Wexler (2003) (which do not explicitly state the immobile results). The higher rate in the active condition than in the passive one is because of the stationarity assumption. In the immobile condition, both the stationarity and rigidity assumptions favour the same percept.

Condition	Experiment	Model
Active	54.3%	79.6%
Immobile	3.6%	17.8%

Because our model implements both the rigidity and stationarity assumptions, they are in competition when the most rigid and most stationary objects do not match. In this experiment, such a mismatch happens in an active condition. The model deals with this kind of contradiction in a way similar to Bayesian fusion (Lebeltel et al., 2004). Other instances of Bayesian fusion are exemplified in the literature (Landy et al., 1995, Ernst and Banks, 2002). The uncertainty, as quantified by the probability distributions, will ensure the optimal balance between the rigidity and stationarity assumptions.

2.6 Effect of Shear on Tilt Perception

Another point that we tested with the Bayesian model is the effect of the shear component of optic flow on SFM performance. The shear angle is the absolute difference between the tilt angle and the direction of the frontal translation. It is called "winding angle" by Cornilleau-Pérès et al. (2002). Psychophysical studies have found that SFM performance in immobile human observers (namely, judgement of tilt) deteriorates drastically as shear increases (Cornilleau-Pérès et al., 2002), but that this deterioration is much less drastic in active observers generating optic flow through their own head movements (Van Boxtel et al., 2003). Examples of minimal and maximal shear in optic flow are shown in Fig. 9. Shear can be parameterized by the *shear angle*, which takes values between 0°, corresponding to no shear, and 90°, corresponding to maximal shear.

We compare model results with experimental findings by Van Boxtel et al. (2003). The experiment involves a monocular observer who is either immobile or moving in a direction perpendicular to the gaze direction (active condition). In both cases, the observer receives the same optic flow. In the active condition, the simulated plane is stationary in an observer-independent reference frame. In the immobile condition, the plane rotates around an axis in the fronto-parallel plane. The observer's task is to report the plane's orientation by aligning a probe so that it appears parallel to the plane.

Figure 10 shows the distribution of absolute tilt errors from the experimental results (Van Boxtel et al., 2003), in both active and immobile conditions, for minimal and maximal shear. First, we see that mean errors increase with increasing shear. However, this effect is much stronger in the immobile condition (where response is almost at the chance level).

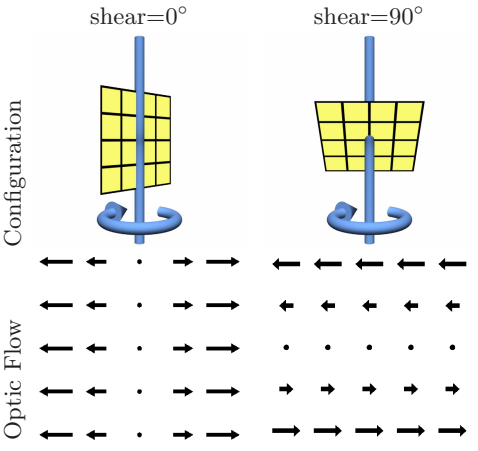


Fig. 9. Illustration of shear in optic flow. Shear can be parameterized by the shear angle, defined as 90° minus the absolute value of the difference between tilt and axis angles. Configurations corresponding to two values of shear angle are shown, 0° (minimum shear) and 90° (maximum). The bottom row shows the optic flow resulting from each configuration.

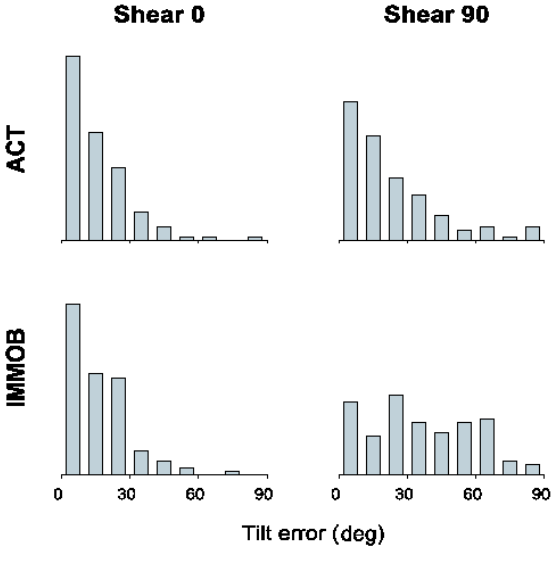


Fig. 10. Tilt error for both active and immobile conditions and shear at 0° and 90°, by Van Boxtel et al. (2003). Depth reversals (much more common in the immobile condition, see Fig. 6) were corrected by using the opposite tilt from the one reported in calculating errors when a reversal occurred; thus, absolute tilt error runs between 0° and 90°.

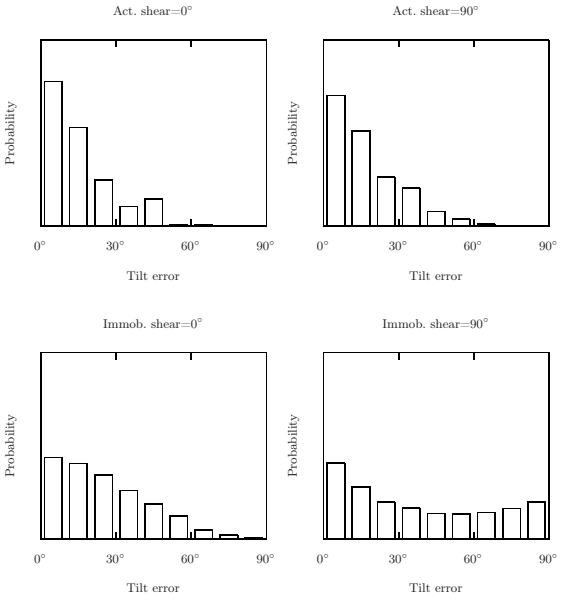


Fig. 11. The effect of shear and observer motion on tilt error, as predicted by the Bayesian model. As in the experimental results (Fig. 10), the mean tilt error is greater for a 90° shear than for 0° shear, and this effect is greater for an immobile observer than an active one.

Figure 11 shows the distribution of absolute tilt errors for the same conditions as given by the model. The variation of the precision between low and high shear is similar to the experimental results.

In the model, the main factor inducing the shear effect is the relative weights of the rotation prior and the translation prior. For a small shear, the absolute motion satisfying the first-order optic flow equations for a large tilt error is composed of a rotation and a translation. For a high shear, a large error corresponds to an absolute motion composed of two rotations with the same velocity. The stationarity assumption states that both the translation and the rotation components of the absolute motion are probably small. However, the constraint on the translation component should be stronger than that on the rotation component to reduce the dispersion of tilt error for small shear. The strength of the shear effect depends on the relative strength of the stationarity constraints on translation and rotation components.

This experiment was conducted using a small field of vision. Our model also predicts a reduced shear effect in a large field of vision, which has actually been found in human observers (Cornilleau-Pérès et al., 2002).

2.7 Influence of Eye Movements on 3D Vision

Using a sinusoidally curved surface that underwent lateral translation while being pursued with the eyes by the subject, Naji and Freeman (2004) found few depth reversals. However, when the same optic flow was presented without pursuit (i.e. with the translation subtracted), depth reversals were prevalent. We simulated a very similar experiment, the only difference being that we used a planar rather than a curved surface. Because planes can undergo depth reversals in the same way as curved surfaces, the main effect of Naji and Freeman, or something very close to it, can be simulated within the framework of our model.

As can be seen in Fig. 3(d) (analogous to condition C of Naji and Freeman (2004)), depth reversals can take place in the pursuit condition. Both solutions undergo the same translation, and one of the solutions additionally undergoes a rotation. In the fixation condition (analogous to condition B of Naji and Freeman (2004)), the same optic flow leads to two solutions undergoing equal and opposite rotations, as shown in Fig. 3(b). Finally, Naji and Freeman (2004) have a third condition (A) where the object translates as in condition C but in which the observers were required to fixate on a stationary point rather than pursue the object.

The rate of depth reversals is calculated from subjects' responses in a depth-order task. Figure 12 shows the experimental results in these three conditions. The graphs show the estimation of the phase with respect to the amplitude of the stimulus. The phase is the analogue of the orientation of the plane in Figs 3(d) and (b), whereas the amplitude stands for the slant of the plane (negative slant being a reversal). We notice that translation (A and C) allows for the disambiguation of the stimulus, whereas rotation exhibits a symmetric behaviour. We notice that the perception is more precise in the pursuit condition (C) than the immobile condition (A).

In comparison, Fig. 13 shows the results of the Bayesian model in the transposed conditions. We can see that the major properties are reproduced, in particular, the broader uncertainty in condition A compared with condition C, as well as the ambiguity in condition B.

Up to now, subjective responses have been limited to the plane orientation. An additional element must be included in the model to account for the 'top-far' responses. This decision was made using a simple Bayesian program. As can be seen in conditions A and C in Fig. 12, the observers exhibited some preference towards a 'top-far' perception. This preference is included as a prior in the Bayesian post-processing. However, it is to be noted that observers seem to have a preference for a 'top-near' perception in condition B.

The results in condition B are the same as in the immobile condition above. The small asymmetry of both top and bottom curves comes from the second-order optic flow that induces a reversal rate strictly less than 50%.

The difference between the model results in conditions A and C comes from the stationarity of the reverse percepts. In condition C, the reverse percept undergoes a greater rotation than in condition A. Therefore, the stationarity assumption assigns a smaller probability to it, hence yielding a smaller reversal rate.

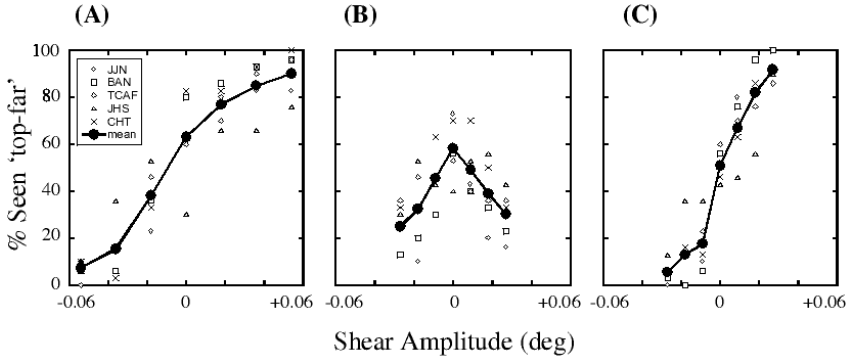


Fig. 12. Rate of 'top-far' perception with respect to the strength of the stimulus (Naji and Freeman, 2004). Condition A corresponds to a translating object without eye pursuit, condition B to a rotating object, and condition C to a translating object with pursuit. Conditions A and B show that translation allows for a disambiguation contrary to passive rotation. Furthermore, a comparison of conditions A and C shows that pursuit of the object leads to better perception.

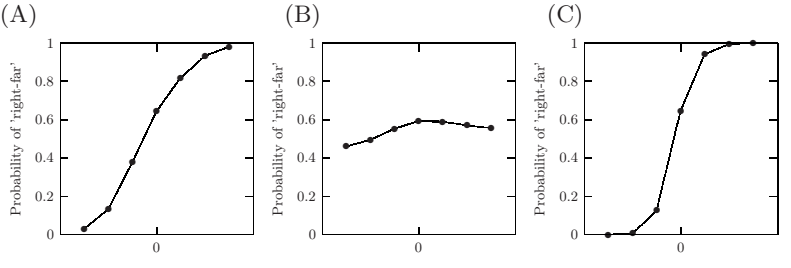


Fig. 13. Results from the model. As for the experimental results, conditions A and C allow for a disambiguation of the stimulus, and condition C is less uncertain than condition A.

3 Discussion

3.1 Probabilistic Expression of Assumptions

A Bayesian model infers the logical consequences of a given set of assumptions with some observations. The inference can occur as soon as a joint probability distribution is defined. Therefore, the modeller must express the assumptions in a Bayesian way.

Each choice in a specification is an assumption. As there are multiple steps in the specification of a joint probability distribution, there are multiple levels of expression. The first choice is the variables and their domain. The variables absent at this step cannot have a meaningful influence in the model. Then, the

joint distribution is decomposed into a product of factors using conditional independencies, which express a lack of relationship between variables and therefore simplify the inference. The last level of expression of assumptions is in the specification of each distribution and its parameters appearing into the decomposition.

Each choice reduces the number of degrees of freedom of the joint distribution. The more drastic restrictions are in the choice of variables and their domains, while the less important are in the choice of the parameters of the distributions. Any reduction can be postponed to a later stage, but the earlier it is done, the more the inference can take advantage of it to simplify the computations.

3.2 Choices in Our Bayesian Model

The issue of specification is therefore to express our assumptions in terms of choice of variable, simplification of the joint distribution using conditional independencies, and choice of parametric forms and parameters.

The first main assumption is that of rigidity, which states that the optic flow more likely to be observed is generated by a plane in relative motion. The parametric space of the optic flow is derived from this assumption. The optic flow is defined by eight parameters. While sufficient for a plane, the optic flow is, in general, more complicated. This means that other eventual components are not relevant variables in our model, and are therefore ignored. It could be interesting to investigate an eventual effect of these components in the human perception of a plane. As far as the model is concerned, such investigation can be studied with additional components in the optic flow variable. Rigidity is also pre-eminent in the choice of the parametric form of the probability distribution over optic flow, given relative motion, position of the plane, and the conditions of observation. We fixed this as a Gaussian distribution. However, it would be possible to evaluate this choice of distribution by measuring the evolution of performance with respect to some additional noise in the stimulus and comparing it with the predicted evolution of the model. The other main assumption of our model is that of stationarity, which states that the motion of the plane is more likely to be small.

The variables chosen to describe the motion are the translation and rotation components along the three axes, according to the experiments chosen as references. This is restrictive in the sense that it does not take into account eventual accelerations and even more complex trajectories. Most reported studies deal with uniform motion; however, investigation of the influence of accelerations in the perception of structure could benefit from the model. The model can be adapted to implement and predict the results of different hypotheses to be compared with experimental results.

The parameters are the last elements of choice in the model. We obtained the results presented above with a single set of parameters. Each experimental result gives information on the exact effect highlighted by the experiment on some parameters. However, each experiment has different optimal parameters; therefore, the final set of parameters chosen results from a trade-off between all the experiments.

3.3 Model Results

The results of the model display some discrepancies with the results of the experiments. For example, for the first experiment described, the reversal rate of the model in a small field is 44.6% compared with 48.8% in the experiment (Cornilleau-Pérès et al., 2002). There are two main reasons for this difference. First, the Bayesian model is a model of an observer. It is not specifically designed to reproduce mean results across observers. Nevertheless, the results of our model are less variable than the results reported between observers (in this case, the minimum reversal rate reported by Cornilleau-Pérès et al. (2002) is around 38%).

As explained above, the set of parameters is the same across all the results of our model. However, there are variations in the precise experimental conditions between the different teams responsible for the measured results. For instance, the rate of reversal measured in a small field of view for an immobile observer by Van Boxtel et al. (2003) is 35%, compared with the 48.8% measured by Cornilleau-Pérès et al. (2002). This can be explained by differences in the protocols that are not taken into account as relevant variables in the Bayesian model. Therefore, as a general rule, the parameters we chose for the Bayesian model are a trade-off between all the results. Accordingly, the results of the model cannot match the experimental results precisely.

The Bayesian model not only accounts for previously reported results but also can be used to make predictions and eventually to propose new experiments. For example, we propose the investigation of the relative influence of stationarity and rigidity in large fields of view. In this case, in an experimental set-up similar to that of Wexler (2003), our model predicts that rigidity will be of greater importance in the perception of second-order optic flow through a diminution of the standard deviations on these components.

Another prediction of the Bayesian model involves the shear effect. In our model, this effect is accounted for by relative weight between rotation and translation components in a small field of vision. Our model predicts a reduced shear effect in large fields of vision, and this has been found in human observers (Cornilleau-Pérès et al., 2002).

References

Cornilleau-Pérès, V., Wexler, M., Droulez, J., Marin, E., Miège, C., Bourdoncle, B.: Visual perception of planar orientation: dominance of static depth cues over motion cues. Vision Research 42, 1403–1412 (2002)

Dijkstra, T., Cornilleau-Pérès, V., Gielen, C., Droulez, J.: Perception of three-dimensional shape from ego- and object-motion: comparison between small- and large-field stimuli. Vision Research 35(4), 453–462 (1995)

Domini, F., Caudek, C.: 3-d structure perceived from dynamic information: a new theory. Trends in Cognitive Sciences 7(10), 444–449 (2003)

Domini, F., Caudek, C.: Perceiving surface slant from deformation of optic flow. J. Exp. Psychol Hum. Percept Perform 25(2), 426–444 (1999)

Ernst, M., Banks, M.: Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415(6870), 429–433 (2002)
Koenderik, J.: Optic flow. Vision Research 26(1), 161–179 (1986)
Landy, M., Maloney, L., Johnston, E., Young, M.: Measurement and modeling of depth cue combination: in defense of weak fusion. Vision Research 35, 389–412 (1995)
Lebeltel, O., Bessière, P., Diard, J., Mazer, E.: Bayesian robot programming. Advanced Robotics 16(1), 49–79 (2004), http://emotion.inrialpes.fr/bibemotion/2004/LBDM04
Longuet-Higgins, H.: The visual ambiguity of a moving plane. In: Proceedings of the Royal Society of London (B, Biological Sciences), vol. 223, pp. 165–175 (1984)
Mayhew, J., Longuet-Higgins, H.: A computational model of binoculard depth perception. Nature 297(5865), 376–378 (1982)
Naji, J., Freeman, T.: Perceiving depth order during pursuit eye movement. Vision Research 44, 3025–3034 (2004)
Rogers, B., Graham, M.: Motion parallax as an independent cue for depth perception. Perception 8, 125–134 (1979)
Rogers, B., Rogers, S.: Visual and nonvisual information disambiguate surfaces specified by motion parallax. Perception and Psychophysics 52, 446–452 (1992)
Todd, J., Bressan, P.: The perception of 3-dimensional affine structure from minimal apparent motion sequences. Perception and Psychophysics 45(5), 419–430 (1990)
Todd, J., Norman, J.: The visual perception of smoothly curved surfaces from minimal apparent motion sequences. Perception and Psychophysics 50(6), 509–523 (1991)
Ullman, S.: The interpretation of visual motion. MIT Press, Cambridge (1979)
Van Boxtel, J., Wexler, M., Droulez, J.: Perception of plane orientation from self-generated and passively observed optic flow. Journal of Vision 3(5), 318–332 (2003), http://journalofvision.org/3/5/1/
Von Helmholtz, H.: Handbuch der Physiologischen Optik. Voss, Hamburg (1867)
Wallach, H., O'Connell, D.: The kinetic depth effect. Journal of Experimental Psychology 45, 205–217 (1953)
Wallach, H., Stanton, J., Becker, D.: The compensation for movement-produced changes in object orientation. Perception and Psychophysics 15, 339–343 (1974)
Wexler, M.: Voluntary head movement and allocentric perception of space. Psychological Science 14, 340–346 (2003)
Wexler, M., Lamouret, I., Droulez, J.: The stationarity hypothesis: an allocentric criterion in visual perception. Vision Research 41, 3023–3037 (2001)
Wexler, M., Panerai, F., Lamouret, I., Droulez, J.: Self-motion and the perception of stationary objects. Nature 409, 85–88 (2001)

A Equations of Optic Flow

Let $\mathcal{P}$ be the object plane, (χ, υ) its depth gradients, $\widetilde{M}$ with coordinates $(\tilde{x}, \tilde{y}, \tilde{z})$ a point of this plane in the 3D reference frame, and M with coordinates (x, y) its projection in the image plane. The equation of the plane is:

$$\tilde{x}\chi + \tilde{y}\upsilon - \tilde{z} = 0. \tag{10}$$

We have the slant of the plane $\sigma = \arctan\sqrt{\chi^2 + \upsilon^2}$ and the tilt $\tau = \arctan\frac{\upsilon}{\chi}$.

Let P be the projection of a 3D point in the image.

$$P: \quad \begin{pmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \end{pmatrix} \mapsto \begin{pmatrix} x = \frac{\tilde{x}}{1-\tilde{z}} \\ y = \frac{\tilde{y}}{1-\tilde{z}} \end{pmatrix}. \tag{11}$$

Let $\mathbf{t} = (t_x, t_y, t_z)$ and $\boldsymbol{\omega} = (\omega_x, \omega_y, \omega_z)$ be respectively the translation and rotation vector of the object plane.

Considering the points as functions of time, we can write:

$$M(t) = P \circ \widetilde{M}(t). \tag{12}$$

Optic flow is the displacement of the points in the image:

$$\boldsymbol{\phi} = \frac{dM}{dt} \tag{13}$$

$$\boldsymbol{\phi} = \frac{dP}{d\widetilde{M}}(\widetilde{M}) \times \frac{d\widetilde{M}}{dt}. \tag{14}$$

$\frac{dP}{dM}$ is the Jacobian of P.

$$\begin{aligned} \frac{dP}{d\widetilde{M}}(\widetilde{M}) &= \begin{pmatrix} \frac{\partial x}{\partial \tilde{x}} & \frac{\partial x}{\partial \tilde{y}} & \frac{\partial x}{\partial \tilde{z}} \\ \frac{\partial y}{\partial \tilde{x}} & \frac{\partial y}{\partial \tilde{y}} & \frac{\partial y}{\partial \tilde{z}} \end{pmatrix} \\ \frac{dP}{d\widetilde{M}}(\widetilde{M}) &= \begin{pmatrix} \frac{1}{1-\tilde{z}} & 0 & \frac{\tilde{x}}{(1-\tilde{z})^2} \\ 0 & \frac{1}{1-\tilde{z}} & \frac{\tilde{y}}{(1-\tilde{z})^2} \end{pmatrix} \end{aligned} \tag{15}$$

The plane $\mathcal{P}$ undergoes translation $\mathbf{t}$ and rotation $\boldsymbol{\omega}$. Therefore the motion $\frac{d\widetilde{M}}{dt}$ of $\widetilde{M}$ is as follows.

$$\begin{aligned} \frac{d\widetilde{M}}{dt} &= \mathbf{t} + \boldsymbol{\omega} \wedge \widetilde{M} \\ &= \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} + \begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix} \wedge \begin{pmatrix} \tilde{x} \\ \tilde{y} \\ \chi\tilde{x} + \upsilon\tilde{y} \end{pmatrix} \\ \frac{d\widetilde{M}}{dt} &= \begin{pmatrix} t_x + \chi\omega_y\tilde{x} + (\upsilon\omega_y - \omega_z)\tilde{y} \\ t_y + (\omega_z - \chi\omega_x)\tilde{x} - \upsilon\omega_x\tilde{y} \\ t_z + \omega_x\tilde{y} - \omega_y\tilde{x} \end{pmatrix} \end{aligned} \tag{16}$$

Substituting 15 and 16 in equation 14, we obtain the following.

$$\begin{aligned} \boldsymbol{\phi} &= \frac{dP}{d\widetilde{M}}(\widetilde{M}) \times \frac{d\widetilde{M}}{dt} \\ &= \begin{pmatrix} \frac{1}{1-\tilde{z}} & 0 & \frac{\tilde{x}}{(1-\tilde{z})^2} \\ 0 & \frac{1}{1-\tilde{z}} & \frac{\tilde{y}}{(1-\tilde{z})^2} \end{pmatrix} \times \begin{pmatrix} t_x + \chi\omega_y\tilde{x} + (\upsilon\omega_y - \omega_z)\tilde{y} \\ t_y + (\omega_z - \chi\omega_x)\tilde{x} - \upsilon\omega_x\tilde{y} \\ t_z + \omega_x\tilde{y} - \omega_y\tilde{x} \end{pmatrix} \\ \boldsymbol{\phi} &= \begin{pmatrix} \frac{t_x + \chi\omega_y\tilde{x} + (\upsilon\omega_y - \omega_z)\tilde{y}}{1-\tilde{z}} + \frac{\tilde{x}}{1-\tilde{z}} \times \frac{t_z + \omega_x\tilde{y} - \omega_y\tilde{x}}{1-\tilde{z}} \\ \frac{t_y + (\omega_z - \chi\omega_x)\tilde{x} - \upsilon\omega_x\tilde{y}}{1-\tilde{z}} + \frac{\tilde{y}}{1-\tilde{z}} \times \frac{t_z + \omega_x\tilde{y} - \omega_y\tilde{x}}{1-\tilde{z}} \end{pmatrix} \end{aligned} \tag{17}$$

By the definition of P (equation 11), $\frac{\tilde{x}}{1-\tilde{z}} = x$, $\frac{\tilde{y}}{1-\tilde{z}} = y$ and $\frac{1}{1-\tilde{z}} = 1+\chi x+\upsilon y$. We can finally rewrite the equation 17 to obtain the equations 18 of the optic flow of a plane, as follows.

$$\boldsymbol{\phi} = \begin{pmatrix} \frac{t_x+\chi\omega_y\tilde{x}+(\upsilon\omega_y-\omega_z)\tilde{y}}{1-\tilde{z}} + \frac{\tilde{x}}{1-\tilde{z}} \times \frac{t_z+\omega_x\tilde{y}-\omega_y\tilde{x}}{1-\tilde{z}} \\ \frac{t_y+(\omega_z-\chi\omega_x)\tilde{x}-\upsilon\omega_x\tilde{y}}{1-\tilde{z}} + \frac{\tilde{y}}{1-\tilde{z}} \times \frac{t_z+\omega_x\tilde{y}-\omega_y\tilde{x}}{1-\tilde{z}} \end{pmatrix}$$

$$\boldsymbol{\phi} = \begin{pmatrix} t_x + x\left[t_z + \chi\left(t_x + \omega_y\right)\right] + y\left[-\omega_z + \upsilon\left(t_x + \omega_y\right)\right] \\ +x^2\left(\chi t_z - \omega_y\right) + xy\left(\upsilon t_z + \omega_x\right) \\ t_y + x\left[\omega_z + \chi\left(t_y - \omega_x\right)\right] + y\left[t_z + \upsilon\left(t_y - \omega_x\right)\right] \\ +xy\left(\chi t_z - \omega_y\right) + y^2\left(\upsilon t_z + \omega_x\right) \end{pmatrix}$$

$$\boldsymbol{\phi} = \boldsymbol{\phi}^0 + \boldsymbol{\phi}^1.{}^t(x,y) + {}^t(x,y).{}^t\boldsymbol{\phi}^2.{}^t(x,y) \tag{18}$$

with:

$$\boldsymbol{\phi}^0 = \begin{pmatrix} t_x \\ t_y \end{pmatrix}$$

$$\boldsymbol{\phi}^1 = \begin{pmatrix} t_z + \chi\left(t_x + \omega_y\right) & -\omega_z + \upsilon\left(t_x + \omega_y\right) \\ \omega_z + \chi\left(t_y - \omega_x\right) & t_z + \upsilon\left(t_y - \omega_x\right) \end{pmatrix}$$

$$\boldsymbol{\phi}^2 = \begin{pmatrix} \chi t_z - \omega_y \\ \upsilon t_z + \omega_x \end{pmatrix}$$