Noname manuscript No.
(will be inserted by the editor)

Francis Colas · Jacques Droulez · Mark Wexler · Pierre Bessière

A unified probabilistic model of the perception of three-dimensional structure from optic flow

the date of receipt and acceptance should be inserted later

Abstract Human observers can perceive the three-dimensional (3-D) structure of their environment using various cues, an important one of which is optic flow. The motion of any point's projection on the retina depends both on the point's movement in space and on its distance from the eye. Therefore, retinal motion can be used to extract the 3-D structure of the environment and the shape of objects, in a process known as *structure-from-motion* (SFM). However, because many combinations of 3-D structure and motion can lead to the same optic flow, SFM is an ill-posed inverse problem. The rigidity hypothesis is a constraint supposed to formally solve the SFM problem and to account for human performance. Recently, however, a number of psychophysical results, with both moving and stationary human observers, have shown that the rigidity hypothesis alone cannot account for human performance in SFM tasks, but no model is known to account for the new results. Here, we construct a Bayesian model of SFM based mainly on one new hypothesis, that of stationarity, coupled with the rigidity hypothesis. The predictions of the model, calculated using a new and powerful methodology called Bayesian programming, account for a wide variety of experimental findings.

F. Colas
INPG – Gravir laboratory, 655 avenue de l'Europe, 38334 Montbonnot, France

F. Colas · P. Bessière
INRIA Rhône-Alpes – Gravir laboratory, 655 avenue de l'Europe, 38334 Montbonnot, France

J. Droulez · M. Wexler
Collège de France – Laboratoire de la Physiologie de la Perception et de l'Action, 11 place Marcelin Berthelot, 75005 Paris, France

J. Droulez · M. Wexler
CNRS – UMR 7152, 11 place Marcelin Berthelot, 75005 Paris, France

P. Bessière
CNRS – UMR 5527, 655 avenue de l'Europe, 38334 Montbonnot, France

1 Introduction

Relative motion between an observer and the objects in a visual scene leads to a deformation of the image on the retina, called optic flow. Optic flow is generated by the 3-D motion of shapes, therefore it contains information, which is relevant to recover the scene geometry. Motion parallax and the kinetic depth effect are special cases of this phenomenon, noticed by Helmholtz (1867), and experimentally quantified by Wallach and O'Connell (1953).

Although it is simple to derive the optic flow corresponding to given 3-D geometry and motion, perception faces the inverse problem, to derive 3-D shape and motion from optic flow. Because an infinite number of combinations of geometry and motion can lead to the same optic flow, SFM is an ill-posed inverse problem.

Formally, the SFM problem can be solved if the *rigidity hypothesis* holds, that is if optic flow is only due to 3-D translations and rotations of a rigid body. In this case, the number of degrees of freedom associated with motion is drastically reduced and both structure and motion can be theroretically recovered from very little optic flow information (Ullman 1979). Several algorithms based on the rigidity hypothesis for special cases, such as planes, have been developed (Mayhew and Longuet-Higgins 1982). Psychophysical results show that human performance on some SFM tasks is at least broadly consistent with predictions based on the rigidity hypothesis (Wallach and O'Connell 1953; Koenderik 1986). However, more recent affine models are based only on local velocity information, rather than on the entire optic flow field, to account for human perception (Todd and Bressan 1990; Todd and Norman 1991); then, even perception of affine properties has been questioned (Domini and Braunstein 1998). It has also been shown that sometimes human perception do not abide to the rigidity hypothesis even if a rigidity interpretation of a stimulus exists (Wexler et al. 2001a).

Most studies of SFM involve an immobile observer experiencing optic flow consistent with moving 3-D objects. However, it is known that SFM is also effective when optic flow is generated by the observer's own head movement about a stationary 3-D scene (Rogers and Graham 1979). Until recently, it has been thought that 3-D shapes perceived in subject-motion SFM are the same as those perceived in object-motion SFM, as long as the optic flow is the same (Wallach et al. 1974; Rogers and Graham 1979). However, in some cases, this turns out to be false: even when optic flow is kept constant, the observer's movement influences perceived 3-D shape (Rogers and Rogers 1992; Dijkstra et al. 1995; Wexler et al. 2001b).

This influence of self-motion on perceived 3-D shape lead to the postulate of a second hypothesis in the interpretation of optic flow, that of *stationarity*: the visual system prefers the solution whose motion is minimal in an observer-independent, allocentric reference frame (Wexler et al. 2001b,a; Wexler 2003). This can be supported by the observation that most of the objects in the visual field are static. While the rigidity hypothesis may be seen as the minimization of *relative* motion between the points of a possible object, the stationarity hypothesis is the minimization of *absolute* object motion in an observer-independent reference frame. Taken separately, neither the stationarity nor the rigidity hypothesis can explain human SFM performance. However, until now, no coherent model has integrated these two hypotheses.

In this article, we present a generic Bayesian model that integrates the stationarity and rigidity hypotheses in the perception of 3-D surfaces from optic flow. The aim is to build a Bayesian model of an observer presented with uncertain stimuli. Then, we instantiate the generic model for the perception of 3-D planar surfaces. This choice is motivated by the availability of data, as well as the complexity of analysis and calculation. This model not only accounts for SFM performance in moving and stationary observers that led to the postulation of the stationarity hypothesis, but also for a number of other, sometimes puzzling, results that have been previously reported. We investigate experiments focusing on the monocular perception of a rotating planar patch with a neutral or non-informative texture. In these experiments, motion was the only cue for plane orientation. However, we look into variations in the experimental conditions involving the motion of the observer's head or eyes, or the plane, or both, as well as the size of the displayed stimulus. Although perception of planes is a special case, it is a very important special case of spatial vision, as the visual world is composed mostly of surfaces, which, if sufficiently regular, can be locally approximated as planes.

In recent years, growing attention has been paid to Bayesian inference, as a common theoretical framework, to understand perceptive skills and multimodal interactions (Weiss et al. 2002; Ernst and Banks 2002; Kersten et al. 2004). In most works, however, the probabilistic reasoning has been limited to simplified forms of Bayes' theorem. One simplification is to use the linear Gaussian assumption, which is not valid in the case of optic flow processing as the different information sources do not combine linearly (see appendix A). The other simplification is to restrict Bayesian models to a combination of prior knowledge and a set of observations. In order to combine several hypotheses, such as rigidity and stationarity, in a mathematically correct form, we found it necessary to put perception models back into a more general Bayesian framework that includes not only observed sensory data and perceived states, but also intermediate variables. Focusing on the SFM problem, we show here that our general Bayesian formalism allows us to express and to test several hypotheses originating from psychophysical experiments, in a very natural and efficient way.

2 Methods

We first present a generic, unified model of perception of the structure of an object from optic flow. Then, we give a precise instantiation for the perception of planes that yields the results presented in section 3.

2.1 Generic model

The generic Bayesian model we propose is the expression of the hypotheses evoked above. The first two are the rigidity (H1) and stationarity (H2) hypotheses. We also assume that the structure of the object is independent of its motion, the motion of the observer and the conditions of observation (H3), and that the conditions of observations are independent of the motions of both the object and the observer (H4).

For sake of simplicity, we have called "rigidity hypothesis" the expression of the relationship between the observed optic flow and the unknown 3-D object structure and motion. In the following, we will describe the object motion relative to the observer as any combination of 3-D rotation and translation. Therefore, we excluded any explicit description of object deformation in motion variable. Any deviation from a strictly rigid object is then entirely defined as a mismatch between the observed optic flow and the optic flow that can be predicted from rigid object transformation. But, of course, there is a natural extension of the present model with explicit non rigid object transformation, adding for instance the first order deformation tensor to the 3-D translation and rotation in the description of object movement. Having restrained the object motion description to isometric transformation does not imply that the perceived movement will be the rigid transformation that explains the best the observed optic flow, since other hypotheses, and particularly the stationarity hypothesis, might induce a strong

deviation from the best rigid solution. The stationarity hypothesis is expressed in the probabilistic relationship between the object's movement with respect to the observer's reference frame and the observeer's movement with respect to the allocentric reference frame. It states that the most probable relative object movement is equal and opposite to the observer's movement. As a consequence, the most probable object movement in the allocentric reference frame is null. In the following, we have not included the object movement in the set of variables, since it can be directly reconstructed by combining the relative object movement and the observer's movement. Rigidity (H1) and stationarity (H2) are the two main hypotheses of our model. As they are expressed in probabilistic forms, none of them could be simultaneously satisfied. The most probable output would be rather the optimal compromise. The last two hypotheses are common assumptions made explicit. Indeed, H3 states that the shape of the object does not influence a priori the motion of the observer, nor the motion of the object itself, nor the conditions by which the object is observed. These conditions of observations can be the size of the image (as in the ollowing example) or the dots density on the object. Hypothesis H4 further adds that the conditions of observation do not influence the object motion and the subject motion. Both these hypotheses reflect the experimental protocols and help reduce the complexity of the model.

We follow the Bayesian programming methodology to specify a model with these hypotheses (Lebeltel et al. 2004). This model uses probabilities to represent and handle the uncertainty faced by an observer. This is a model of what an observer can deduce from the limited information of optic flow.

As a summary, we have a set of hypotheses related to the experiment and a methodology for the specification of a Bayesian model. At each step of the methodology, we extract the relevant knowledge from the hypotheses.

From relevant information to variables The unified model is based on relevant variables common to all instances of structure-from-motion perception. Additional variables can be used to comply with specific experimental conditions.

In this context, we propose a model that takes into account: (i) the observed optic flow (noted $\mathbf{\Phi}$), (ii) the 3-D structure of the object (noted Θ), (iii) the motion of the object (noted X) in the observer's reference frame, (iv) the motion of the observer in the allocentric reference frame (noted $\mathbf{M}$), and (v) the general viewing conditions as defined by the experimental protocol (noted Λ).

Due to our rigidity hypothesis, we restrict the form of relative motion of the object and self-motion to isometrics transformations of the 3-D space.

As this is a generic model, these are formal variables. In the next section, presenting the instantiation of this generic model for the case of a moving plane, these variables will be given actual mathematical expressions.

From dependencies to decomposition At the core of a Bayesian model lies the joint probability distribution over all its variables. This joint distribution follows from the assumptions 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*. Bayesian programming methodology includes making the formal simplifications of the decomposition before actually dealing with the specification of each factor.

The structure and relative motion of the object are sufficient to define the optic flow of an object. Therefore, the absolute self-motion is unnecessary for the optic flow. This corresponds to the following mathematical simplification:

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

The stationarity hypothesis (H2) states that object motion is most likely to be small in the allocentric reference frame. This defines a constraint on $P(X \mid \mathbf{M})$ (see next section); therefore we use Bayes' rule to write:

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

The application of Bayes' rule to $P(\mathbf{M}\ X)$ can lead also to $P(X)\ P(\mathbf{M} \mid X)$ but the stationarity hypothesis will be a simpler to express with equation 2.

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

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

The last hypothesis (H4) states the independence between the motions and the general viewing conditions:

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

Finally, using Bayes' rule, we can write :

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

Putting together equations 5, 3, 4, 2, and 1, we obtain the decomposition, shown in equation 6, that is the structural expression of our hypotheses.

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

The decomposition states the lack of relation between some of our generic variables. In this case, the structure of the object, the conditions of observation, and self-motion are independent; relative motion only depends on self-motion, due to the stationarity hypothesis; and optic flow does not depend on self-motion, but only on relative motion, structure of the object, and conditions of observation.

From physical and physiological laws to distributions The decomposition only state whether there is a relation between variables. In order to get a usable expression for the joint distribution, these relations have to be defined. This is done by specifying each of the probability distributions that appear as factors of the decomposition of the joint distribution.

The first factor, $P(\Theta)$, is the prior on the structure of the object. As we build the model of perception by an observer, it represents what this observer expects before any observation. It can be an uninformative prior or it can reflect some bias in perception, in favor of more common shapes.

In the same way, $P(\mathbf{M})$ and $P(\Lambda)$ represent respectively the expectation by an observer of her or his own motion, and of the conditions of observation. If we consider that the model has an exact knowledge of them (as will be the case later in this article), this probability distribution is simplified in the final inference and thus can be left unspecified.

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, this means a Gaussian distribution. A dirac distribution set to the opposite of self-motion would mean absolute certainty of a non-moving object in the absolute reference frame, and would therefore rule out any interpretation of the stimulus involving a moving object.

The last factor in decomposition 6 is the distribution of optic flow, given the structure of the object, the relative motion between the object and the observer, and the conditions of observation, $P(\mathbf{\Phi} \mid \Theta\ X\ \Lambda)$. The rigidity hypothesis states that the optic flow is generated by a rigid object in motion. Therefore, we specify this factor saying that the most probable optic flow is the theoretical optic flow of the object in this particular configuration, given this particular motion as can be computed by standard optics calculations. It can be interpreted as the optic flow formation process, relaxed by a non-null probability of a different optic flow for a given situation. A dirac distribution on the exact theoretical flow would rule out any non rigid interpretation of a given optic flow.

Formalized questions A probabilistic question is the 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 the rules of probability calculus.

The participants of the experiments have to answer a unique value to solve the task, instead of a probability distribution. Without any cost function to specify the decision process, we sample the distribution computed to answer the probabilitistic question. As a consequence, over repeated trials, the distribution of answers of our model approaches to the probability distribution from which they are sampled. Error distributions can be computed directly from these distributions, without resorting to any stochastic process.

The precise question we ask to solve the SFM issue is the probability of the object structure or shape, given the optic flow, the self-motion, and the general conditions of observation written as $P(\Theta \mid \boldsymbol{\phi}\ \mathbf{m}\ \lambda)$.[1] This question is answered by the following expression, that results from Bayes' rule, marginalization rule and use of the decomposition (expression 6):

$$\begin{aligned} & P(\Theta \mid \boldsymbol{\phi}\ \mathbf{m}\ \lambda) \\ &= \frac{\sum_{x \in X} P(\Theta)P(\lambda)P(\mathbf{m})P(x \mid \mathbf{m})P(\boldsymbol{\phi} \mid \Theta\ x\ \lambda)}{P(\boldsymbol{\phi}\ \mathbf{m}\ \lambda)} \\ &\propto P(\Theta) \sum_{x \in X} P(x \mid \mathbf{m}) P(\boldsymbol{\phi} \mid \Theta\ x\ \lambda). \end{aligned} \tag{7}$$

This is essentially the problem we solved to obtain the results shown later in this article. Given observations of optic flow and self-motion, this distribution represents knowledge about the structure of the object (including its relative position with respect to the observer) that one can infer from our hypotheses. The observations do not need to be noiseless. If the added uncertainty (for example on optic flow) is compatible the probability distributions (in this case the variance on $P(\mathbf{\Phi} \mid \Theta\ X\ \Lambda)$), the model will behave essentially the same as with clean input.

Furthermore, the same probabilistic model can be used to answer other questions. For example, one may be interested in the estimation of self-motion from optic flow: $P(\mathbf{M} \mid \boldsymbol{\phi}\ \lambda)$. This question can be used to study vection, where optic flow induces the sensation of self-motion, and the direction of perceived self-motion, called heading. For this question, Bayesian inference with the same model gives the following expression:

$$\begin{aligned} & P(\mathbf{M} \mid \boldsymbol{\phi}\ \lambda) \\ &= \frac{\sum_{x \in X, \theta \in \Theta} P(\theta)P(\lambda)P(\mathbf{M})P(x \mid \mathbf{M})P(\boldsymbol{\phi} \mid \theta\ x\ \lambda)}{P(\boldsymbol{\phi}\ \lambda)} \\ &\propto P(\mathbf{M}) \sum_{x \in X, \theta \in \Theta} P(\theta) P(x \mid \mathbf{M}) P(\boldsymbol{\phi} \mid \theta\ x\ \lambda). \end{aligned} \tag{8}$$

2.2 The case of a moving dotted plane

The generic model is a template, which must be adapted to account for particular experiments. The remaining of this paper will focus on the perception of a moving planar object. This allows for a simpler actual model than with other kind of surfaces while still exhibiting interesting properties of the ambiguity of perception. In this section, we present the instantiated model for this particular case that we use to generate the results presented in the next section.

[1] We use an uppercase letter for a variable and lowercase for the instantiation of a variable with a particular value.

Variables For this model, we need only consider instantaneous variables, as the experiments deal with short stimuli without large change during the course of its presentation. However, the model can be adapted to time-varying variables with exactly the same instantaneous structure of dependency. The structure Θ of the object is reduced to the position and orientation of the plane. As one point of the plane is already known (the fixation point)[2], only two orientation parameters are needed to parametrize the structure of the object. For practical reasons, we use the depth gradients along the transversal and vertical axes. If we call x, y, and z the coordinates of a point of the plane along the transversal, vertical, and sagittal axes respectively, then the structure Θ is the pair $(\chi, \upsilon) = \left(\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}\right)$.

Self-motion $\mathbf{M}$ is a set of translation and rotation velocities of the observer, chosen along the transversal, vertical, and sagittal axes. Likewise, relative motion is decomposed into its rotation and translation components, $\mathbf{\Omega}$ and $\mathbf{T}$ respectively.

In the case of planar objects, the optic flow is entirely specified by eight components (see appendix for details), namely the two velocity components at the origin ($\mathbf{\Phi}^0$), the four first-order derivatives of the velocity field at the origin ($\mathbf{\Phi}^1$), and the two independent components of the second-order derivatives of the velocity field at the origin ($\mathbf{\Phi}^2$) (Longuet-Higgins 1984).

Finally, we restrain the viewing condition parameters to the most critical one, the size of the field of view.

Distributions The prior on plane orientation $P(\Theta)$ is chosen to be the least informative, so as not to bias the inference. This corresponds to a prior invariant to arbitrary rotation of the plane. Others prior can be chosen based on ecological arguments, for example in favour of the horizontal plane. However, lacking precise experimental data, we opted for an unbiased distribution.

For SFM question $P(\Theta \mid \boldsymbol{\phi}\ \mathbf{m}\ \lambda)$, both self-motion, $\mathbf{m}$, and the size of field, λ, are known. The posterior distribution does not depend on the priors on variables $\mathbf{M}$ and Λ; therefore, these prior distributions do not need to be specified, as can be seen in expression 7.

As for the expression of stationarity, the distribution of relative motion given self-motion $P(X \mid \mathbf{M})$ yields the most probable relative motion as equal-and-opposite to self-motion, corresponding to no absolute motion. To this end, we choose a Gaussian distribution centered on such relative motion. Indeed, the Gaussian is the least informative distribution, given the mean and the uncertainty of the distribution. It also corresponds to the Gibbs distribution with kinetic energy. Choosing a least informative distribution ensures that we do not put additional constraints into the model that do not appear in our list of hypotheses.

[2] By convention the distance between the fixation point and the observer is taken as the unit of distance. This way the scale issue disappears.

Likewise, the distribution of optic flow, given the relative motion and orientation of the plane and the size of the field of view, is an expression of the rigidity hypothesis. We chose a Gaussian distribution centered on the theoretical values of the eight components (see expression in appendix A). The field of view is assumed to change the variance of the second-order components ($\mathbf{\Phi}^2$). Indeed, in a smaller field of view, second-order components are much more difficult to extract than in large field of view compared to first-order components.

Implementation Although the specified distributions are either Gaussian or uniform, the SFM question has no analytical solution because of the intrinsic nonlinearities of the optic flow equations (see appendix). Quantitative simulations are then performed by computing the exact inference on discretized variables. Table 1 gives the details of the domains of the variables. ranges (minimum, maximum and number of samples in between) and dimensionality of each component of Θ (top row), of the relative rotation (second row), of the relative translation (third row), and of the size of the field of view (bottom row). Other variables do not need to be discretized as their values are known for the inference.

On the other hand, some of the distributions in our decomposition involve parameters. This is the case with the Gaussians on relative motion and optic flow, whose parameters are shown in table 2. We use a single set of parameters for all the results of the following section. Each covariance matrix was determined accordingly to reasonnable values for all the experiments then fitted one by one with a local search against global results. Therefore the parameters are a trade-off between the different experiments. The calculations are lead using the ProBT inference engine (Lebeltel et al. 2004).

Distribution parameters
$\sigma_{\mathbf{T}} = 0.3 * Id_{3\times3}$ in $m.s^{-1}$
$\sigma_{\mathbf{\Omega}} = 1.2 * Id_{3\times3}$ in $rad.s^{-1}$
$\sigma_{\Phi^0} = 1.0 * Id_{2\times2}$ in $m.s^{-1}$
$\sigma_{\Phi^1} = 0.025 * Id_{4\times4}$ in s^{-1}
$\sigma_{\Phi^2 \mid \lambda=SF} = 5.0 * Id_{2\times2}$ in $m^{-1}.s^{-1}$
$\sigma_{\Phi^2 \mid \lambda=LF} = 0.2 * Id_{2\times2}$ in $m^{-1}.s^{-1}$

Table 2: Covariance matrices of 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 view and order 2 optic flow in a large field of view.

3 Results

There are numerous sources of ambiguity in the perception of optic flow. Figures 1 and 6 show five kinds of situations of motion of the object or the observer that

Variable	Symbol	Minimum Value	Maximum Value	Number of Values by Dimension	Dimension
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
Size of field	Λ	$0.015\ sr$	$1.05\ sr$	2	1

Table 1: Domain (minimum, maximum and number of samples in between) and dimensionality of each component of Θ (top row), of the relative rotation (second row), of the relative translation (third row), and of the size of the field of view (bottom row).

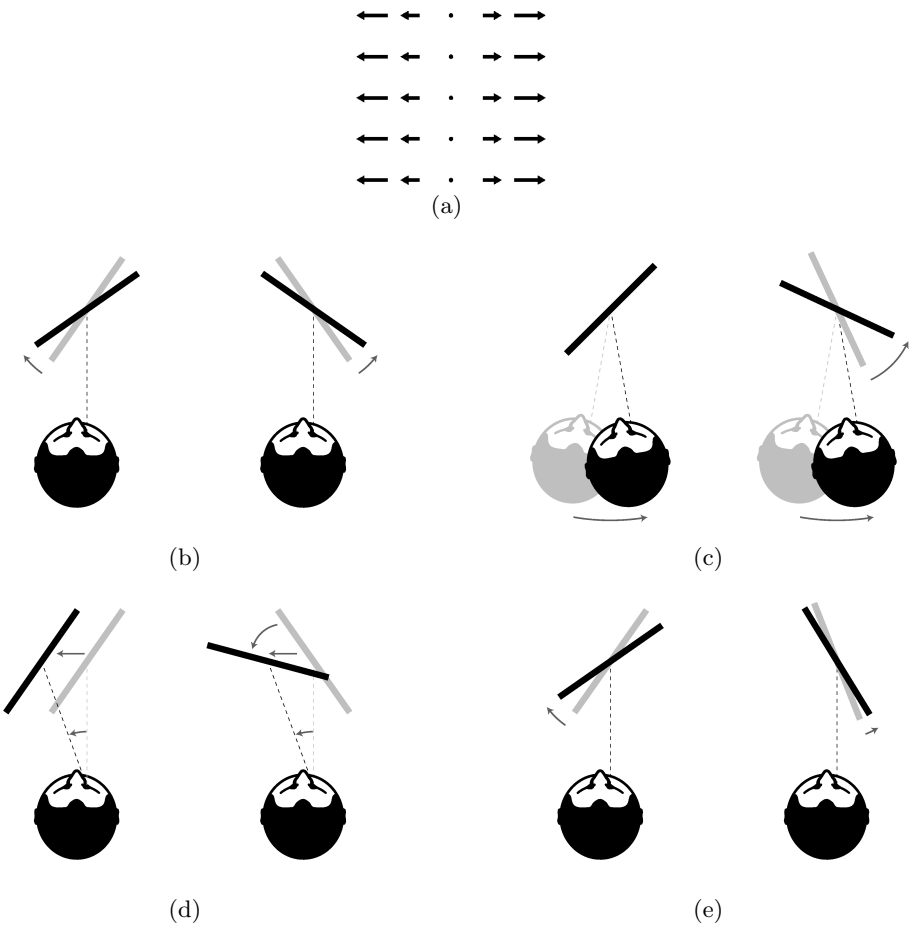

Fig. 1: Some ambiguities in first-order optic flow that have been used in the studies cited. (a) An example of an optic flow field that presents a number of ambiguities: all configurations shown in this figure lead to this flow. (b) The two configurations, which differ by simultaneous reversals of relative depth and 3-D motion, both yield the optic flow shown in (a). This ambiguity is called depth reversal. (c) Depth reversals can also occur for moving observers. The 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 but also a rotation. (e) Ambiguity between slant and rotation speed: a larger slant coupled with a slower rotation speed may give the same optic flow as a lower slant together with a faster rotation.

generate approximately the same optic flow. They have been studied in detail by six sets of psychophysics experiments previously reported. We show that the Bayesian model compares to human performance in various conditions of motion of the plane, voluntary motion of the observer, and size of field of view.

3.1 Depth reversal

Depth reversal is a well-known effect in monocular vision: many depth cues are ambiguous about the sign of relative depth (cf. the Necker cube). In SFM the simplest instance of this ambiguity is the observation of a rotat-

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

Table 3: Influence of the size of field of view on reversal rate. Both the experiments (Cornilleau-Pérès et al. 2002) and the Bayesian model exhibit less reversal percept in a large field of view.

ing plane through a small opening. In this case, there is an ambiguity on the tilt and direction of rotation, as illustrated in figure 1(b). The extrinsic orientation of a plane in 3-D space is often parametrized by two angles; slant and tilt. Slant is the angle, in 3-D 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. In this case, a depth reversal is characterized by the perception of tilt and rotation of the plane in the opposite direction at the same time (see figure 1(b)). However it has been shown that this ambiguity does not hold for a large field of view (Dijkstra et al. 1995). We will investigate this simple effect as the first example of our model.

The experiment we use as a reference has been described by Cornilleau-Pérès et al. (2002). In this experiment, the stationary participant observes a planar patch in rotation about a fronto-parallel axis (the plane is painted with a uniform random dot texture). After the presentation of the stimulus, the observer is asked to estimate the orientation of the planar patch by aligning a probe to it. Two field-of-view sizes were compared: a large field with a 60° aperture angle and a small field with an 8° aperture angle.

Cornilleau-Pérès et al. (2002) report the results in terms of the rate of tilt reversals. A tilt reversal is defined to occur when absolute error in the estimation of the tilt angle is greater than 90°. The reversal rate can be considered a measure of the ambiguity, as illustrated in figure 1(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 small field of view to below 5% in large field of view.

Our 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 figure 2). Ambiguity in the optic flow interpretation, such as illustrated in figure 1, results in a multimodal probability distribution. To compare the reversal rate reported by Cornilleau-Pérès et al. (2002) with model output, we computed the sum of probabilities corresponding to tilt errors greater than 90° (see table 3).

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

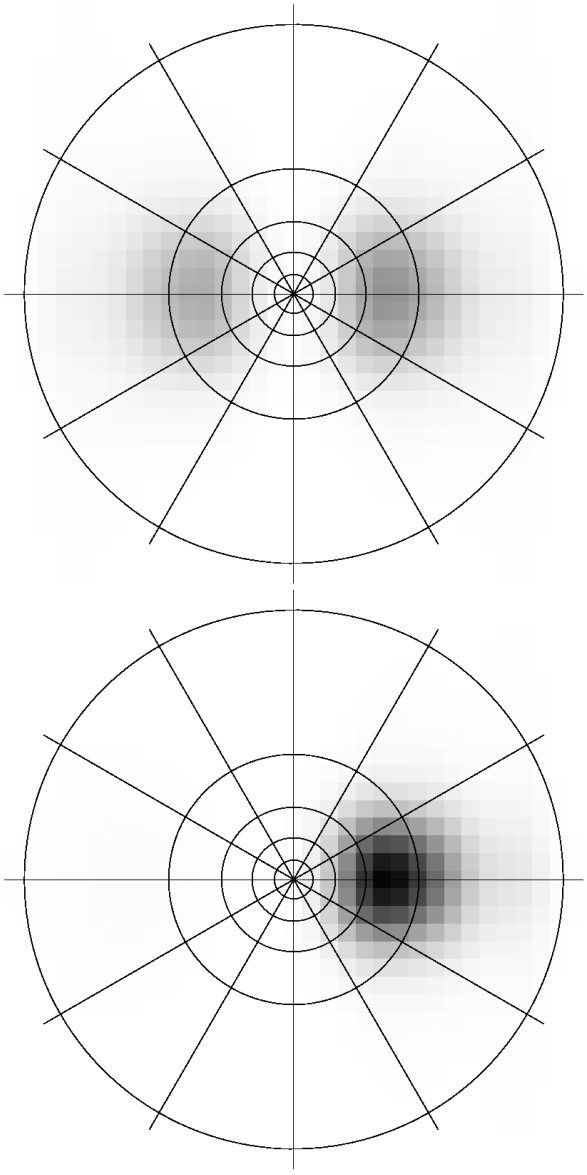

Fig. 2: Examples of probability distributions on the orientation of a plane. The polar angle is the tilt of the plane, the radius is the tangent of the slant angle, and the color stands for the probability. A darker color represents a higher probability. The peaks represent the most likely percepts, 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 and the lower panel displays a low reversal rate.

In the Bayesian model, the standard deviation over the second-order optic flow is smaller in a large field than in a small field of view. Therefore the influence of second-order optic flow is greater in a large field of view than in a small field.

Qualitatively, insofar as this uncertainty is greater in a small field, the probability of reversal will always be higher in a small field than in a large field. Figure 3 shows the quantitative evolution of the reversal rate in the model as a function of this parameter.

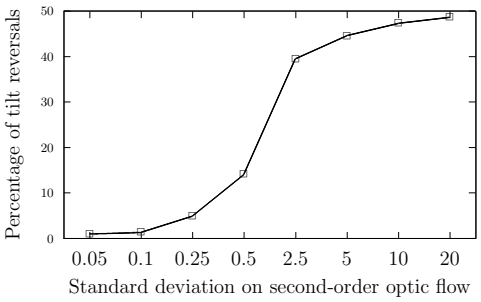

Fig. 3: Influence of the uncertainty of second-order optic flow on the prediction of reversal rate in the Bayesian model. A small field of view leads to a greater uncertainty, and hence to more reversals.

3.2 Depth reversals in moving and immobile observers

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 in moving and immobile observers, while keeping optic flow the same in the two self-motion conditions. Thus, actively generated optic flow can lead to a different perception of 3-D 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 figure 1a leads, in the immobile observer, to total ambiguity between the solutions shown in figure 1b, and therefore a depth reversal rate of up to 50% for a small field of view. In the moving observer (figure 1c), on the other hand, the ambiguity is lifted in favor of the solution that is most stationary in an observer-independent reference frame (the left solution in figure 1c).

The experimental data used as a reference is taken from van Boxtel et al. (2003), in which the perception of the same optic flow is compared in active and immobile conditions (figure 4), in a small field of view. The experimental results clearly 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 subject'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°.

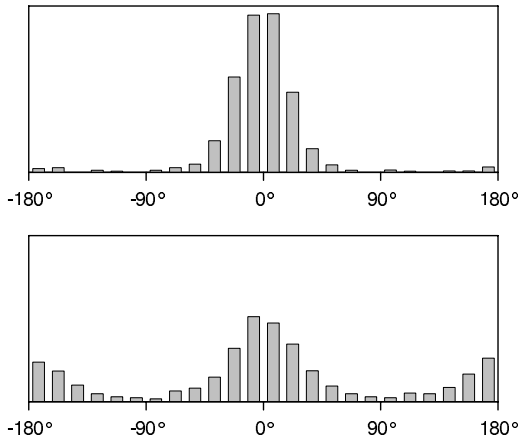

Fig. 4: Distributions of error in tilt angle for both active (top) and immobile (bottom) conditions, by van Boxtel et al. (2003). The results show depth reversals in the immobile condition and its almost complete disappearance in the active condition.

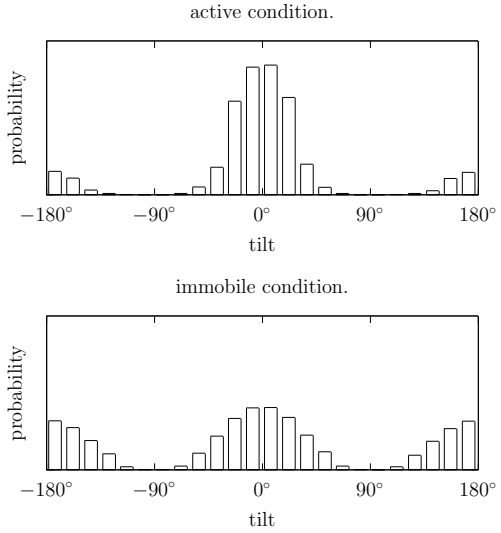

Fig. 5: Probability distributions of tilt errors in active and immobile conditions. As in the experimental results shown in figure 4, the ambiguity drastically diminishes in the active condition.

Figure 5 shows the results of our model in the same two conditions. They were computed in a similar way than in the previous experiment: we applied a variable change to the posterior distribution on structure computed by our model in order to compute the posterior probability distribution over tilt errors. We notice that the bimodality in the immobile condition is similar to the

experimental results, and the decrease of reversals in the active condition. In the Bayesian model, the bimodality shown above is derived from the symmetry of the 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 hypothesis in our model. In the immobile condition, the simulated and depth-reversed planes have the same speed, as depicted in figure 1(b); only the direction of motion changes. In the active condition, however, the simulated plane is stationary in an observer-independent reference frame, whereas the depth-reversed plane has high velocity 1(c). Therefore, the stationarity hypothesis, as implemented in the model, insures that the reversed plane is less probable, because it corresponds to a higher velocity in an observer-independent reference frame.

3.3 Ambiguity between slant and speed

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

The experimental data we consider are taken from Domini and Caudek (1999). The experimental conditions involve a static monocular observer in a small field of view. 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 velocity, where the second is twice that of the first.

The experimental results, by 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, 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 and 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. As before, these were computed by a variable change on the posterior distribution of the model to compute the posterior distribution on the tangent of the slant angle. Then we computed the mean of this new distribution, like in the experimental results. 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 with increasing 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 hypothesis (see Methods for details).

It is noted that the values of perceived slant for the model are smaller than those of 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.[3]

	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

Table 4: Mean perceived tangent of slant as a function of simulated slant tangent and angular speed for 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.

The slant/speed ambiguity results from ambiguities in first-order optic flow. Indeed, in both situations (small slant, high speed compared to large slant, low speed) the optic flow is the same up to the order one as shown in figure 1(e), and only the second-order optic flow could disambiguate the stimulus. These results confirm the low weighing of the second-order components of optic flow in a small field of view. This low weighing is due to the uncertainty attached to the distribution over the second-order optic flow.

First-order optic flow can be partially described by a parameter called *def*, the product of the tangent of the slant and angular speed (Domini and Caudek 2003).[4] Therefore slant and speed cannot be recovered individually from first-order optic flow. Domini and

[3] One possible influence for this difference is the size of the field of view, which is larger in this experiment than for the others.

[4] 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.

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 $P(\mathbf{\Phi}^1 \mid \mathbf{\Omega}\ \Theta)$. The norm of 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.

3.4 Ambiguity of translation in depth

Another symmetry or ambiguity of first-order optic flow is shown in figure 6. A rotation in depth generates the same (first-order) optic flow as a translation 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 depth translation, that is, if the observer moves toward the surface, or if the surface moves toward the observer (see figure 6). These results can be summarized by stating that there is a strong bias toward perceiving the solution that minimizes motion in an observer-independent reference frame. Thus, these results provide further support for the stationarity hypothesis. However, the observer's percepts are also, by and large, in agreement with the rigidity hypothesis. Therefore, they provide a useful testing ground for our model, which incorporates both the stationarity and rigidity hypotheses.

In the psychophysical studies, two conditions are tested: in the active condition, the observer moves his head in depth; in the immobile condition, the observer remains still but experiences the same optic flow as in a previous active trial (Wexler et al. 2001a; Wexler 2003).[5] In the active condition, the optic flow is generated by a plane rotating in depth, where the distance to the observer is fixed (the plane's center therefore undergoes depth translation as well). Therefore, in the active condition 6(d), the rigidity hypothesis favours the simulated plane, while the stationarity hypothesis favours the alternative solution.[6] In the immobile condition, on the other hand, both the rigidity and stationarity hypotheses favour the simulated plane.

[5] Other conditions, involving conflict between the observer's motor command and self-motion, were also tested (Wexler 2003), and found to lead to different response distributions. More precisely, when there is a mismatch between motor command and self-motion, the performance of the observers are similar to unvoluntary motion and significantly different from voluntary motion accurately performed. The model would need an additional variable to tackle this mismatch condition.

[6] The reason why the rigidity hypothesis favours the simulated plane rather than the alternative solution is that the symmetry of figure 6 only holds for first-order optic flow. The second-order terms break the symmetry, and lead to nonrigidity of the alternative solution.

Both experimental results and model results are presented figure 7. Recall that optic flow is the same in the active and immobile conditions; only the observers' motion differs. Providing that only first-order optic flow components are available, the rigidity hypothesis alone would predict equally low rates for the alternative solution 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 nonrigid solution.

As explained above, the discrepancy between the actual values of the experimental results and the model are due to the unique parameter set used for all six experiments. More precisely, different groups of participants already exhibit differences in their results. Compare, for instance, the top left histogram in figure 7 with the bottom left histogram in figure 9. Both correspond to the same conditions but the results are numerically different. Priors in our model can be adjusted to better fit some results at the expense of other experiments.

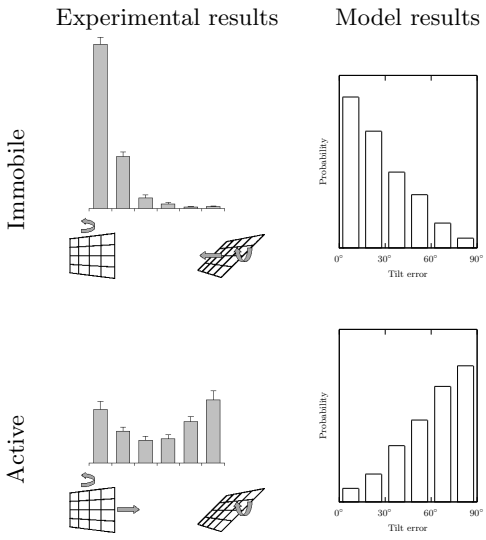

Fig. 7: Distributions of the absolute value of the difference between perceived orientation and rigid solution. The left column shows the experimental results by Wexler (2003) and the right column shows the results from our model, computed by variable change on the posterior distribution. The top row shows results for immobile observers and the bottom row shows results for active observers. These results show that both for the experimental results and the model, perception for an immobile observer will favors rigid and stationary solutions (left bars). In active conditions both results show a higher probability of perception of non-rigid and stationary solutions (right bars). Note that the preference for stationarity of the model is more intense than in the experimental results. This is due to the trade-offs in the choice of a common parameter set for all the experiments.

Because our model implements both the rigidity and stationarity hypotheses, they are in competition when

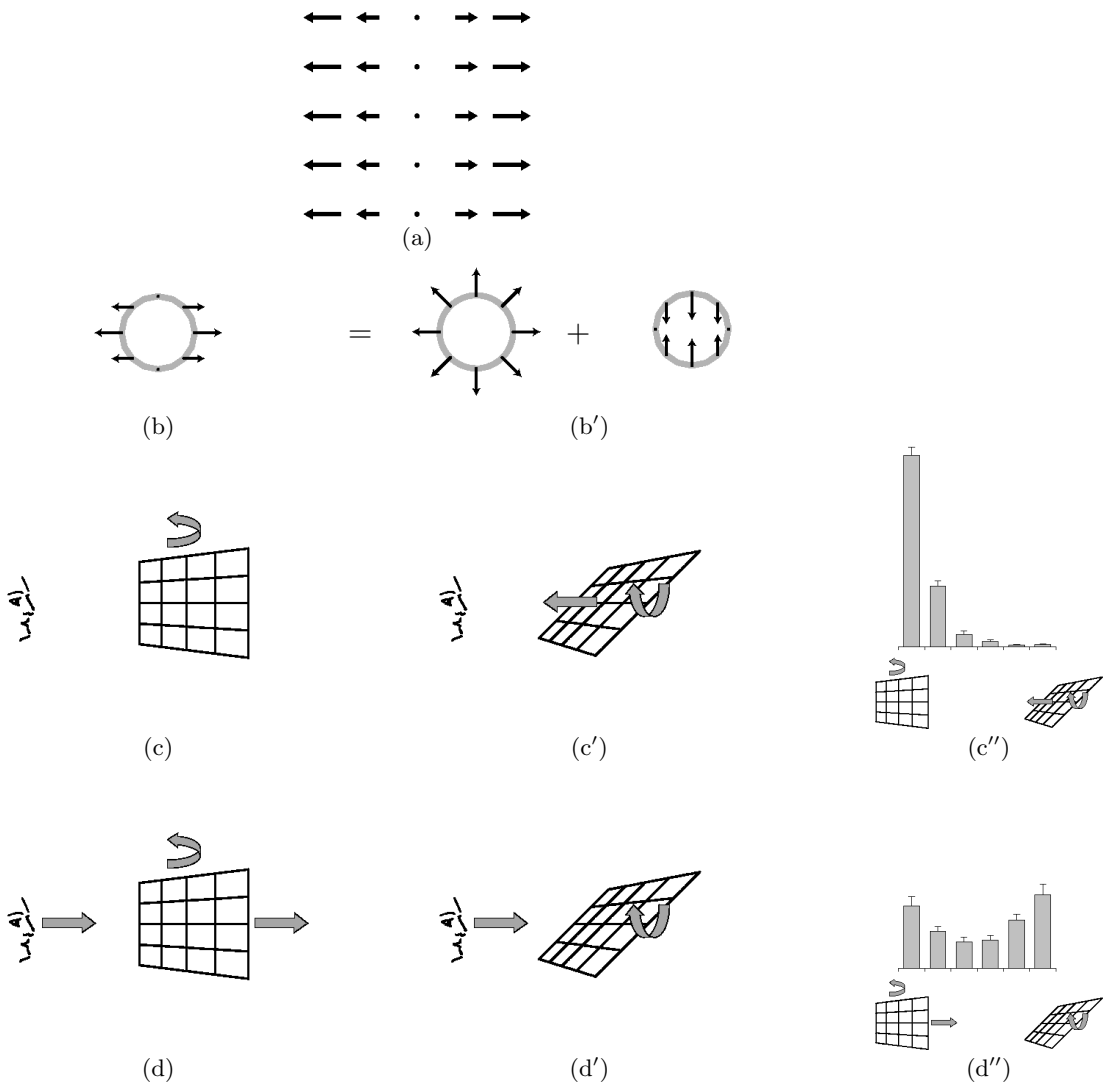

Fig. 6: Illustration of the effect of head motion on the perception of 3-D structures (Wexler et al. 2001a; Wexler 2003). (a) An ambiguous 2D optic flow field that can have different 3-D interpretations, discovered by J. Droulez (cf fig. 1(a)). The arrows represent the motion of projections of points in 3-D space on the retina. It is fairly easy to see that the 3-D 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 center of (d′) that is stationary in space, while (d) translates at the same speed as the observer. (c″), (d″) Data by Wexler (2003) show respectively the frequencies of the absolute value of the difference between perceived orientation and orientation of solution (c) and (d) for stationary (c″) and moving (d″) observers. The bars on the left correspond respectively to solutions (c) and (d), and the bars on the right to solutions (c′) and (d′). Although optic flow is the same in the two cases, perceptions of 3-D structure are very different, showing the effect of the observer's action.

the most rigid and most stationary objects do not match. In this experiment, such a mismatch happens in the active condition. Wexler et al. (2001a) define a rigidity measure and use its symmetry to account for non rigid responses. This model only relies on a sensible rigidity measure which can be the probability as in the present paper. In our model, we can additionally deal with this kind of contradiction in a way that is 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 balance

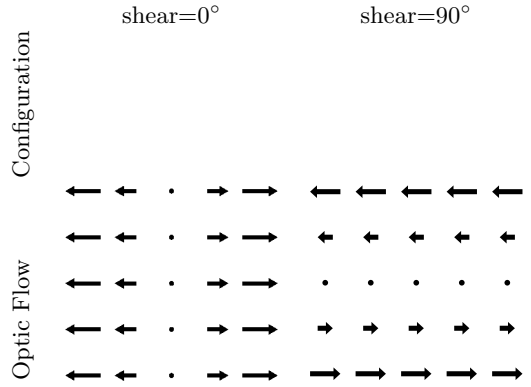

Fig. 8: Illustration of shear in optic flow. Shear can be parametrized by the shear angle, defined as 90° minus the absolute value of the difference between tilt and axis angles. Configurations corresponding to two values of shear angle are shown; 0° (minimum shear) and 90° (maximum). The bottom row shows the optic flow resulting from each configuration.

between the rigidity and stationarity hypotheses. More precisely, both rigidity and stationarity hypotheses are simultaneously maximized by the maximization of the product of the probability distributions reflecting each of those hypotheses, that is $P(\mathbf{\Phi} \mid \Theta\ X\ \Lambda)$ for rigidity and $P(X \mid \mathbf{M})$ for stationarity.

3.5 The effect of shear on SFM

Another point 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 figure 8. Shear can be parametrized by the *shear angle* (which takes values between 0°, corresponding to no shear, and 90°, corresponding to maximal shear).

We compared model results to 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 direction gaze (active condition). In the two conditions, 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 about 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 9 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. We can see that mean errors increase with increasing shear. However, this effect is much stronger in the immobile condition (where response is almost at chance level for highest shear) than in the active condition.

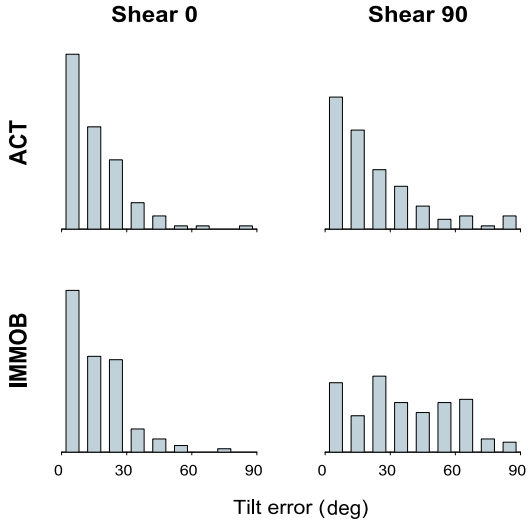

Fig. 9: Tilt error for both active and immobile conditions and shear 0° and 90°, by van Boxtel et al. (2003). Tilt reversals (much more common in the immobile condition, see figure 4) were corrected by using the opposite tilt from the one reported in calculating errors, when an reversal occurred; thus, absolute tilt error runs between 0° and 90°.

Figure 10 shows the distribution of absolute tilt errors for the same conditions as given by the model. As usual, these were computed with a variable change from the posterior distribution computed by our model to the distributions on absolute tilt errors shown. 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 strength of the rotation prior and the translation prior. Indeed, for a small shear, the absolute motion that satisfies 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 hypothesis states that both the translation and the rotation components of the absolute motion are probably small. Therefore solutions corresponding to large error will have their probability reduced by the probability of the object enduring a given rotation and translation, or two rotations for respectively a small or a large shear. If the strength (or, more precisely, the inverse variance of the Gaussian distribution) of the constraint on the translation components is higher

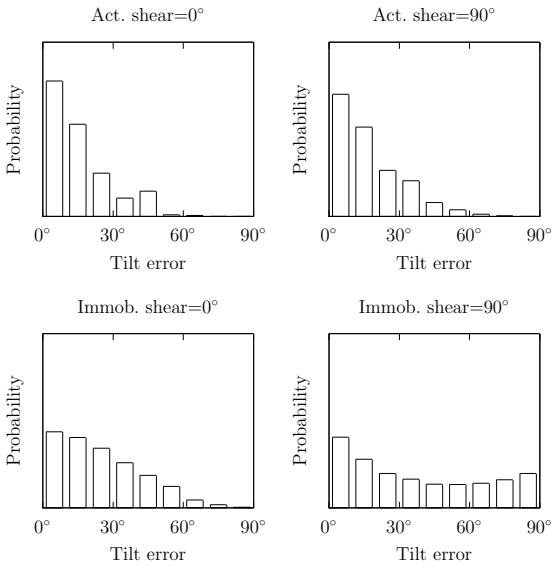

Fig. 10: The effect of shear and observer motion on tilt error, as predicted by the Bayesian model. As in the experimental results (figure 9), the mean tilt error is greater for a 90° shear than for 0° and this effect is greater for an immobile observer than an active one.

than on the rotation components, the probability of a experiencing large errors on tilt will be smaller for a small shear than for a large shear. That is the condition, in the model, to reproduce the shear effect. The strength of this effect depends on the relative strength of the constraints on translation and rotation components: the larger the difference between variance on rotation and on translation, the clearer the effect of the shear on the dispersion of tilt angle perception.

3.6 Influence of eye movements on 3-D 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 substracted), depth reversals were prevalent. We simulated a very similar experiment, with 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 found by Naji and Freeman, or something very close it, can be simulated within the framework of our model.

As can be seen in figure 1d (analogous to condition C by 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 by Naji and Freeman (2004)), the same optic flow leads to two solutions undergoing equal-and-opposite rotations, as shown in figure 1b. 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 11 shows the experimental results of these three conditions. The graphs show the breakdown of the estimation of the phase of the sinusoidal shape (either 'top-far' or 'top-near') with respect to the amplitude of the stimulus. The phase is the analog of the orientation of the plane in figures 1(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 behavior. We notice that the perception is more precise in the pursuit condition (C) than the immobile condition (A).

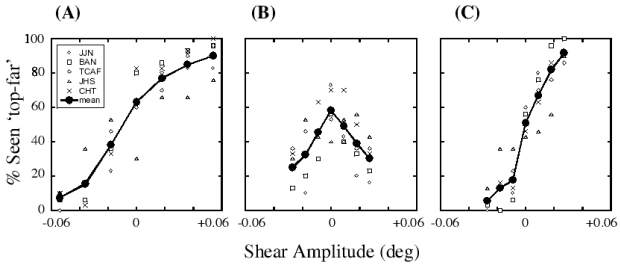

Fig. 11: Rate of 'top-far' perception with respect to the strength of the stimulus (Naji and Freeman 2004). Condition A corresponds to a translating object without eye pursuit; condition B to a rotating object and condition C to a translating object with pursuit. Conditions A and B show that translation allows for a disambiguation, contrary to passive rotation. Furthermore, the comparison of conditions A and C shows that pursuit of the object leads to better perception.

In comparison, figure 12 shows the results of the Bayesian model in the transposed conditions. We can see the major properties are reproduced, in particular the broader uncertainty in condition A compared to condition C, as well as the ambiguity in condition B.

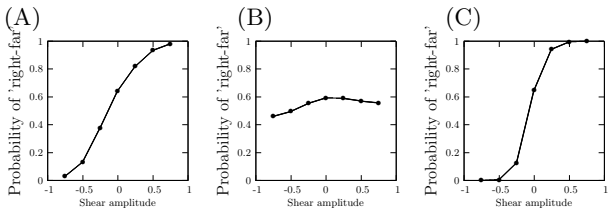

Fig. 12: Results from the model. As for the experimental results, conditions A and C allow for disambiguation of the stimulus, and condition C is less uncertain than condition A.

Until now for the model, subjectives responses were evaluations of the values of plane orientation which can be computed directly from the posterior distribution on structure. For this experiment, an additional element has to be included in the model in order to account for the 'top-far' responses. This was done as a post-processing of the posterior distribution using a simple Bayesian program. As can be seen in conditions A and C in figure 11, the observers exhibited some preference toward 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 those 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 hypothesis assigns it a smaller probability, hence yielding a smaller reversal rate.

4 Discussion

4.1 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 modeler has to express the assumptions in a Bayesian way.

The expression of the assumptions of a Bayesian model can occur at multiple levels, corresponding to the steps of specification of the joint probability distribution. The first level is the choice of the variables and their domain. Variables ruled out at this step cannot have any influence in the model. One step further, the joint probability distribution over the chosen variables is decomposed into a product of factors by the way of conditional independencies. These express a lack of relationship between variables and therefore reduce the complexity of the inference. The final level of expression of assumptions is in the choice of each distribution involved in the decomposition, along with their eventual parameters.

Each choice is a reduction in the degrees of freedom of the joint distribution. The more drastic restrictions are in the choice of the variables and their domain while the less important are in the choice of the parameters of the distribution. 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.

4.2 Choices in our Bayesian model

Designing a Bayesian model is therefore choosing the level of specification to express each of the assumptions.

The first main hypothesis 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 hypothesis. The optic flow is defined by eight parameters. While sufficient in the case of 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 involved in the decomposition with the independancy between optic flow and self-motion conditionally to the knowledge of the structure of the object and the relative motion of the object. Finally, rigidity is preeminent 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 evolution of performance with respect to some additional noise in the stimulus and comparing it to the predicted evolution of the model.

The other main hypothesis 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 optic flow are restricted to instantaneous measurements of displacement of the dots and those for the motion of the object 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 trajectory. 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 handle series of observations and more complex motion, allowing to looking into the results of different hypotheses that can be compared to 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, the optimal parameters for each experiment are different; therefore, the final set of parameters chosen results from a trade-off between all the experiments.

4.3 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 than the variability 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 48.8% measured by Cornilleau-Pérès et al. (2002). This can be explained by differences in the protocol 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. This way, the results of the model cannot precisely match those of the experimental results.

The Bayesian model not only accounts for previously reported results but can also be used to make predictions and eventually 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 setup 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 standard deviation 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 view. Our model predicts a reduced shear effect in large fields of view, and this has been found in human observers (Cornilleau-Pérès et al. 2002).

4.4 Conclusion

In this article, we have presented a generic Bayesian model to integrate both stationarity and rigidity hypotheses for the perception of 3-D surfaces from optic flow. We have detailed the instantiation of such a model to tackle the exemplary case of the perception of the perception of a plane. We presented the results of our model compared with six experimental results from the literature.

The rigidity and stationarity hypotheses are implemented by conditional independencies and probability distributions. In this way, the resulting model could account for many aspects of the perception of an planar object from optic flow. In a more general manner, we think that Bayesian modelling can prove useful to handle the inherent uncertainties of perception.

Acknowledgements This work was supported by the European Project BIBA IST-2001-32115.

A Optic flow equations

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

$$\tilde{x}\chi + \tilde{y}\upsilon - \tilde{z} = 0. \tag{9}$$

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

Let $\mathit{\Pi}$ be the projection of a 3-D point in the image:

$$\mathit{\Pi} : \quad \begin{pmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \end{pmatrix} \mapsto \begin{pmatrix} x = \frac{\tilde{x}}{1-\tilde{z}} \\ y = \frac{\tilde{y}}{1-\tilde{z}} \end{pmatrix}. \tag{10}$$

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

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

$$A(t) = \mathit{\Pi} \circ \widetilde{A}(t). \tag{11}$$

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

$$\boldsymbol{\phi} = \frac{dA}{dt} \tag{12}$$

$$\boldsymbol{\phi} = \frac{d\mathit{\Pi}}{d\widetilde{A}}(\widetilde{A}) \times \frac{d\widetilde{A}}{dt}. \tag{13}$$

$\frac{d\mathit{\Pi}}{d\widetilde{A}}$ is the Jacobian of $\mathit{\Pi}$:

$$\frac{d\mathit{\Pi}}{d\widetilde{A}}(\widetilde{A}) = \begin{pmatrix} \frac{\partial x}{\partial \tilde{x}} & \frac{\partial x}{\partial \tilde{y}} & \frac{\partial x}{\partial \tilde{z}} \\ \frac{\partial y}{\partial \tilde{x}} & \frac{\partial y}{\partial \tilde{y}} & \frac{\partial y}{\partial \tilde{z}} \end{pmatrix}$$

$$\frac{d\mathit{\Pi}}{d\widetilde{A}}(\widetilde{A}) = \begin{pmatrix} \frac{1}{1-\tilde{z}} & 0 & \frac{\tilde{x}}{(1-\tilde{z})^2} \\ 0 & \frac{1}{1-\tilde{z}} & \frac{\tilde{y}}{(1-\tilde{z})^2} \end{pmatrix} \tag{14}$$

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

$$\begin{aligned} \frac{d\widetilde{A}}{dt} &= \mathbf{t} + \boldsymbol{\omega} \wedge \widetilde{A} \\ &= \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} + \begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix} \wedge \begin{pmatrix} \tilde{x} \\ \tilde{y} \\ \chi\tilde{x} + \upsilon\tilde{y} \end{pmatrix} \\ \frac{d\widetilde{A}}{dt} &= \begin{pmatrix} t_x + \chi\omega_y\tilde{x} + (\upsilon\omega_y - \omega_z)\tilde{y} \\ t_y + (\omega_z - \chi\omega_x)\tilde{x} - \upsilon\omega_x\tilde{y} \\ t_z + \omega_x\tilde{y} - \omega_y\tilde{x} \end{pmatrix} \end{aligned} \tag{15}$$

Substituting 14 and 15 in equation 13, we get:

$$\boldsymbol{\phi} = \frac{d\mathit{\Pi}}{d\widetilde{A}}(\widetilde{A}) \times \frac{d\widetilde{A}}{dt}$$

$$= \begin{pmatrix} \frac{1}{1-\tilde{z}} & 0 & \frac{\tilde{x}}{(1-\tilde{z})^2} \\ 0 & \frac{1}{1-\tilde{z}} & \frac{\tilde{y}}{(1-\tilde{z})^2} \end{pmatrix} \times \begin{pmatrix} t_x + \chi\omega_y\tilde{x} + (\upsilon\omega_y - \omega_z)\tilde{y} \\ t_y + (\omega_z - \chi\omega_x)\tilde{x} - \upsilon\omega_x\tilde{y} \\ t_z + \omega_x\tilde{y} - \omega_y\tilde{x} \end{pmatrix}$$

$$\boldsymbol{\phi} = \begin{pmatrix} \frac{t_x+\chi\omega_y\tilde{x}+(\upsilon\omega_y-\omega_z)\tilde{y}}{1-\tilde{z}} + \frac{\tilde{x}}{1-\tilde{z}} \times \frac{t_z+\omega_x\tilde{y}-\omega_y\tilde{x}}{1-\tilde{z}} \\ \frac{t_y+(\omega_z-\chi\omega_x)\tilde{x}-\upsilon\omega_x\tilde{y}}{1-\tilde{z}} + \frac{\tilde{y}}{1-\tilde{z}} \times \frac{t_z+\omega_x\tilde{y}-\omega_y\tilde{x}}{1-\tilde{z}} \end{pmatrix} \quad (16)$$

By definition of Π (equation 10), $\frac{\tilde{x}}{1-\tilde{z}} = x$, $\frac{\tilde{y}}{1-\tilde{z}} = y$ and $\frac{1}{1-\tilde{z}} = 1 + \chi x + \upsilon y$. We can finally rewrite the equation 16 to get the equations 17 of the optic flow of a plane:

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

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

$$\boldsymbol{\phi} = \boldsymbol{\phi}^0 + \boldsymbol{\phi}^1 .^t(x,y) +^t (x,y).^t\boldsymbol{\phi}^2 .^t(x,y) \quad (17)$$

with:

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

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

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

References

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

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

F. Domini and M. Braunstein. Recovery of 3-D structure from motion is neither Euclidean nor affine. *J Exp Psychol Hum Percept Perform.*, 24(4):1273–1295, 1998.

F. Domini and C. Caudek. 3-d structure perceived from dynamic information: a new theory. *Trends in Cognitive Sciences*, 7(10):444–9, 2003.

F. Domini and C. Caudek. Perceiving surface slant from deformation of optic flow. *J Exp Psychol Hum Percept Perform.*, 25(2):426–44, 1999.

M. Ernst and M. Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415 (6870):429–33, 2002.

H. Helmholtz, von. *Handbuch der Physiologischen Optik.* Hamburg: Voss, 1867.

D. Kersten, P. Mamassian, and A. Yuille. Object perception as bayesian inference. *annu Rev Psychol.*, 55:271–304, 2004.

J. Koenderik. Optic flow. *Vision Research*, 26(1):161–179, 1986.

M. Landy, L. Maloney, E. Johnston, and M. Young. Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Research*, 35:389–412, 1995.

O. Lebeltel, P. Bessière, J. Diard, and E. Mazer. Bayesian robot programming. *Advanced Robotics*, 16(1):49–79, 2004. URL `http://emotion.inrialpes.fr/bibemotion/2004/LBDM04/`.

H. Longuet-Higgins. The visual ambiguity of a moving plane. *Proceedings of the Royal Society of London (B, Biological Sciences)*, 223:165–175, 1984.

J. Mayhew and H. Longuet-Higgins. A computational model of binocular depth perception. *Nature*, 297(5865):376–8, 1982.

J. Naji and T. Freeman. Perceiving depth order during pursuit eye movement. *Vision Research*, 44:3025–34, 2004.

B. Rogers and M. Graham. Motion parallax as an independent cue for depth perception. *Perception*, 8:125–134, 1979.

B. Rogers and S. Rogers. Visual and nonvisual information disambiguate surfaces specified by motion parallax. *Perception and Psychophysics*, 52:446–452, 1992.

J. Todd and P. Bressan. The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Perception and Psychophysics*, 45(5):419–30, 1990.

J. Todd and J. Norman. The visual perception of smoothly curved surfaces from minimal apparent motion sequences. *Perception and Psychophysics*, 50(6):509–23, 1991.

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

J. van Boxtel, M. Wexler, and J. Droulez. Perception of plane orientation from self-generated and passively observed optic flow. *Journal of Vision*, 3(5):318–332, 2003. URL `http://journalofvision.org/3/5/1/`.

H. Wallach and D. O'Connell. The kinetic depth effect. *Journal of Experimental Psychology*, 45:205–217, 1953.

H. Wallach, J. Stanton, and D. Becker. The compensation for movement-produced changes in object orientation. *Perception and Psychophysics*, 15:339–343, 1974.

Y. Weiss, E. Simoncelli, and E. Adelson. Motion illusions as optimal percepts. *Nature Neuroscience*, 5(6):508–10, 2002.

M. Wexler. Voluntary head movement and allocentric perception of space. *Psychological Science*, 14:340–346, 2003.

M. Wexler, I. Lamouret, and J. Droulez. The stationarity hypothesis: an allocentric criterion in visual perception. *Vision Research*, 41:3023–3037, 2001a.

M. Wexler, F. Panerai, I. Lamouret, and J. Droulez. Self-motion and the perception of stationary objects. *Nature*, 409:85–88, 2001b.