MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY
and
CENTER FOR BIOLOGICAL AND COMPUTATIONAL LEARNING
DEPARTMENT OF BRAIN AND COGNITIVE SCIENCES

A.I. Memo No. 1624 February, 1998
C.B.C.L. Paper No. 158

Slow and Smooth: a Bayesian theory for the combination of local motion signals in human vision

Yair Weiss and Edward H. Adelson
Dept. of Brain and Cognitive Sciences
MIT E10-120, Cambridge, MA 02139, USA
yweiss@psyche.mit.edu

Abstract

In order to estimate the motion of an object, the visual system needs to combine multiple local measurements, each of which carries some degree of ambiguity. We present a model of motion perception whereby measurements from different image regions are combined according to a Bayesian estimator — the estimated motion maximizes the posterior probability assuming a prior favoring slow and smooth velocities. In reviewing a large number of previously published phenomena we find that the Bayesian estimator predicts a wide range of psychophysical results. This suggests that the seemingly complex set of illusions arise from a single computational strategy that is optimal under reasonable assumptions.

This report describes research done at the Center for Biological and Computational Learning and the Department of Brain and Cognitive Sciences of the Massachusetts Institute of Technology. Support for the Center is provided in part by a grant from the National Science Foundation under contract ASC–9217041. The work was also supported by NEI R01 EY11005 to E. H. Adelson

1 Introduction

Estimating motion in scenes containing multiple, complex motions remains a difficult problem for computer vision systems, yet is performed effortlessly by human observers. Motion analysis in such scenes imposes conflicting demands on the design of a vision system [5]. The inherent ambiguity of local motion signals means that local computations cannot provide enough information to obtain a correct estimate. Thus the system must *integrate* many local measurements. On the other hand, the fact that there are multiple motions means that global computations are likely to mix together measurements derived from different motions. Thus the system also must *segment* the local measurements.

In this paper we are concerned with the first part of the problem, the integration of multiple constraints. Even if we know the scene contains only a single object, estimating that motion is nontrivial. This difficulty arises from the ambiguity of individual velocity measurements which may give only a partial constraint on the unknown motion [39] , i.e. the "aperture problem", [13, 2, 17]. To solve this problem, most models assume a two stage scheme whereby local readings are first computed, and then integrated in a second stage to produce velocity estimates. Psychophysical [2, 20, 42] and neurophysiological [20, 29] findings are consistent with such a model.

The nature of the integration scheme used in the second stage remains, however, controversial. This is true even for the simple, widely studied "plaid" stimulus in which two oriented gratings translate rigidly in the image plane (figure 1a). Due to the aperture problem, only the component of velocity normal to the orientation of the grating can be estimated, and hence each grating motion is consistent with an infinite number of possible velocities, a constraint line in velocity space (figure 1b). When each grating is viewed in isolation, subjects typically perceive the normal velocity (shown by arrows in figure 1b). Yet when the two gratings are presented simultaneously subjects often perceive them moving coherently and ascribe a single motion to the plaid pattern [2, 39].

Adelson and Movshon (1982) distinguished between three methods to estimate this "pattern motion" – Intersection of Constraints (IOC), Vector Average (VA) and blob tracking. Intersection of Constraints (IOC) finds the single translation vector that is consistent with the information at both gratings. Graphically, this can be thought of as finding the point in velocity space that lies at the intersection of both constraint lines (circle in figure 1b). Vector Average (VA) combines the two normal velocities by taking their average. Graphically this corresponds to finding the point in velocity space that lies halfway in between the two normal velocities (square in figure 1b). Blob tracking makes use of the motion of the intersections [8, 19] which contain unambiguous information indicating the pattern velocity. For plaid patterns blob tracking and IOC give identical predictions — they would both predict veridical perception.

The wealth of experimental results on the perception of motion in plaids reveals a surprisingly complex picture. Perceived pattern motion is sometimes veridical (consistent with IOC or feature tracking) and at other times significantly biased towards the VA direction. The degree of bias is influenced by factors including orientation of the gratings [45, 4, 7], contrast [35], presentation time [45] and foveal location [45].

Thus even for the restricted case of plaid stimuli, neither of the three models suggested above can by themselves explain the range of percepts. Instead, one needs to assume that human motion perception is based on at least two separate mechanisms — a "2D motion" mechanism that estimates veridical motion and a crude "1D motion" mechanism that is at times biased away from the veridical motion. Many investigators have proposed that two separate motion mechanisms exist and that these are later combined [30, 15, 19, 3].

As an example of a two mechanism explanation, consider the Wilson et al. (92) model of perceived direction of sine wave plaids. The perceived motion is assumed to be the average of two motion estimates one obtained by a "Fourier" pathway and the other by a "non-Fourier" pathway. The "Fourier" pathway calculates the normal motions of the two components while the "non-Fourier" pathway calculates motion energy on a squared and filtered version of the pattern.

Both pathways use vector average to calculate their motion estimates, but the inclusion of the "non-Fourier" pathway causes the estimate to be more veridical. Wilson et al. have shown that their model may predict biased or veridical estimates of direction depending on the parameters of the stimulus. The change in model prediction with stimulus parameters arises from the fact that the two mechanisms operate in separate regimes. Thus since plaids move in the vector average at short durations and not at long durations, it was assumed that the "non-Fourier" mechanism is delayed relative to the "Fourier" pathway. Since plaids move more veridically in the fovea than in the periphery, the model non-Fourier responses were divided by two in the periphery.

The danger of such an explanation is that practically any psychophysical result on perceived direction can be accommodated - by assuming that the "2D" mechanism operates when the motion is veridical, and does not operate whenever the motion is biased. For example, Alais et al (1994) favor a 2D "blob tracking" explanation for perceived direction of plaids. The fact that some plaids exhibit large biases in perceived direction while others do not is attributed to the fact that some plaids contain "optimal blobs" while others contain "suboptimal blobs" [3]. Although the data may require these types of post-hoc explanations, we would prefer a more principled explanation in terms of a single mechanism.

Evidence that the complex set of experimental results

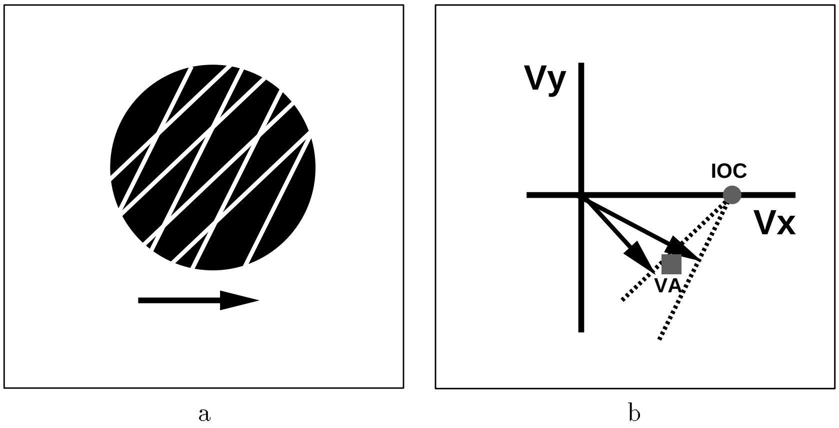


Figure 1: **a.** Two gratings translating in the image plane give a "plaid" pattern. **b.** Due to the aperture problem, the measurements for a single grating are consistent with a family of motions all lying on a constraint line in velocity space. Intersection of Constraints (IOC) finds the single velocity consistent with both sources of information. Vector Averaging (VA) takes the average of the two normal velocities. Experimental evidence for both types of combination rules has been found.

on plaids may indeed be explained using a single principled mechanism comes from the work of Heeger and Simoncelli [11, 33, 32, 34]. Their model consisted of a bank of spatiotemporal filters, whose outputs were pooled to form velocity tuned units. The population of velocity units represented an optimal Bayesian estimate of the local velocity, assuming a prior probability favoring slow speeds. Their model worked directly on the raw image data and could be used to calculate the local velocity for any image sequence. In general, their model predicted a velocity close to the veridical velocity of the stimulus, but under certain conditions (e.g. low contrast, small angular separation) predicted velocities that were biased towards the vector average. They showed that these conditions for biased perception were consistent with data from human observers.

The controversy over the integration scheme used to estimate the translation of plaids may obscure the fact that they are hardly representative of the range of motions the visual system needs to analyze. A model of integration of local constraints in human vision should also account for perception of more complex motions than rigid 2D translation in the image plane. As an example, consider the perception of circles and derived figures in rotation (figure 2). When a "fat" ellipse , with aspect ratio close to unity, rotates in the image plane, it is perceived as deforming nonrigidly [21, 40, 22]. However, when a "narrow" ellipse, with aspect ratio far from unity, rotates in the image plane, the motion is perceived veridically [40].

Unfortunately, the models surveyed above for the perception of plaids can not be directly applied to explain this percept. These models estimate a single velocity vector rather than a spatially varying velocity field. An elegant explanation was offered by Hildreth (1983) using a very different style of model. She explained this and other motion "illusions" of smooth contours with a model that minimizes the variation of the perceived velocity field along the contour. She showed that for a rigid body with explicit features, her model will always give the physically "correct" motion field, but for smooth contours the estimate may be wrong. In the cases when the estimate was physically "wrong", it qualitatively agreed with human percepts of the same stimuli. Grzywacz and Yuille (1991) used a modified definition of smoothness to explain the misperception of smooth contours undergoing rigid translation [23, 24].

Thus the question of how the visual system integrates multiple local motion constraints has not a single answer in the existing literature but rather a multitude of answers. Each of the models proposed can successfully explain a subset of the rich experimental data.

In this paper we propose a single Bayesian model for motion integration and show that it can account for a wide range of percepts. We show that seemingly unconnected phenomena in human vision – from bias towards vector average in plaids to perceived nonrigidity in ellipses may arise from an optimal Bayesian estimation strategy in human vision.

2 Intuition — Bayesian motion perception

In order to obtain intuition about how Bayesian motion perception works, this section describes the construction of an overly simplified Bayesian motion estimator. As we discuss at the end of this section, this restricted model *can not* account for the range of phenomena we are interested in explaining. However, understanding the restricted model may help understand the more general Bayesian model.

While the Bayesian approach to perception has re-

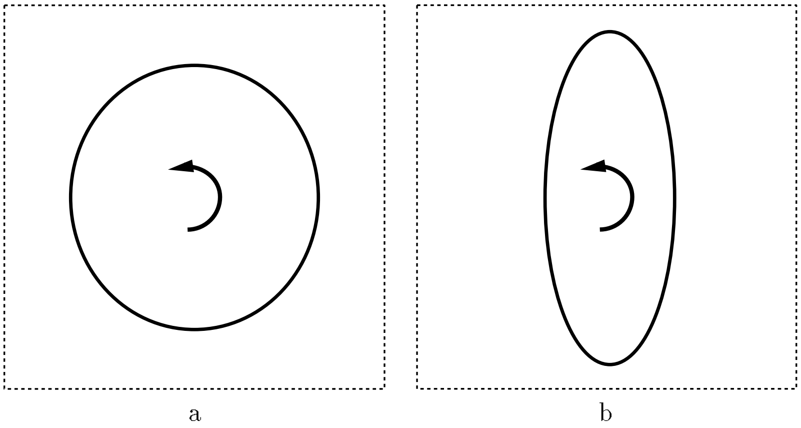


Figure 2: **a.** a "fat" ellipse rotating rigidly in the image plane appears to deform nonrigidly. **b.** a "narrow" ellipse rotating rigidly in the image plane appears to rotate rigidly.

cently been used by a number of researchers (see e.g. [14]), different authors may mean different things when they refer to the visual system as Bayesian. Here we refer to two aspects of Bayesian inference - (1) that different measurements are combined while taking into account their degree of certainty and (2) that measurements are combined together with prior knowledge to arrive at an estimate.

To illustrate this definition, consider an observer who is trying to estimate the temperature outside her house. She sends out two messengers who perform measurements and report back to her. One messenger reports that the temperature is 80 degrees and attaches a high degree of certainty to his measurement, while the second messenger reports that the temperature is 60 with a low degree of certainty. The observer herself, without making any measurements, has prior knowledge that the temperature this time of the year is typically around 90 degrees. According to our definition, there are two ways in which the observer can be a non Bayesian. First, by ignoring the certainty of the two messengers and giving equal weight to the two estimates. Second, by ignoring her prior knowledge and using only the two measurements.

In order to perform Bayesian inference the observer needs to formalize her prior knowledge as a probability distribution and to ask both messengers to report probability distributions as well — the likelihoods of their evidence given a temperature. Denote by θ the unknown temperature, and E_a, E_b the evidence considered by the two messengers. The task of the Bayesian observer is to calculate the posterior probability of any temperature value given both sources of evidence:

$$P(\theta|E_a, E_b) \tag{1}$$

Using Bayes rule, this can be rewritten:

$$P(\theta|E_a, E_b) = kP(\theta)P(E_a, E_b|\theta) \tag{2}$$

where k is a normalizing constant that is independent of θ. Note that the right hand side of equation 2 requires knowing the joint probability of the evidence of the two messengers. Typically, neither of the two messengers would know this probability, as it requires some knowledge of the amount of information shared between them. A simplifying assumption is that the two messengers consider conditionally independent sources of evidence, in which case equation 2 simplifies into:

$$P(\theta|E_a, E_b) = kP(\theta)P(E_a|\theta)P(E_b|\theta) \tag{3}$$

Equation 3 expresses the *posterior* probability of the temperature as a product of the *prior* probability and the *likelihoods*. The *Maximum a posteriore (MAP)* estimate is the one that maximizes the posterior probability.

If the likelihoods and the prior probability are Gaussian distributions, the MAP estimate has a very simple form — it reduces to a weighted average of the two estimates and the prior where the weights are inversely proportional to the variances. Formally, assume $P(E_a|\theta)$ is a Gaussian with mean μ_a and variance V_a, $P(E_b|\theta)$ is a Gaussian with mean μ_b and variance V_b, and the prior $P(\theta)$ is a Gaussian with mean μ_p and variance V_p. Then θ^*, the MAP estimate is given by:

$$\theta^* = \frac{\frac{1}{V_a}\mu_a + \frac{1}{V_b}\mu_b + \frac{1}{V_p}\mu_p}{\frac{1}{V_a} + \frac{1}{V_b} + \frac{1}{V_p}} \tag{4}$$

Equation 4 illustrates the two properties of a Bayesian estimator — the two likelihoods are combined with a prior and all quantities are weighted by their uncertainty.

Motion perception can be considered in analogous terms. Suppose the observer is trying to estimate the velocity of a translating pattern. Different image locations give local readings of the motion with varying degrees of uncertainty and the observer also has some prior probability over the possible velocity. In a Bayesian estimation procedure, the observer would use the local readings in

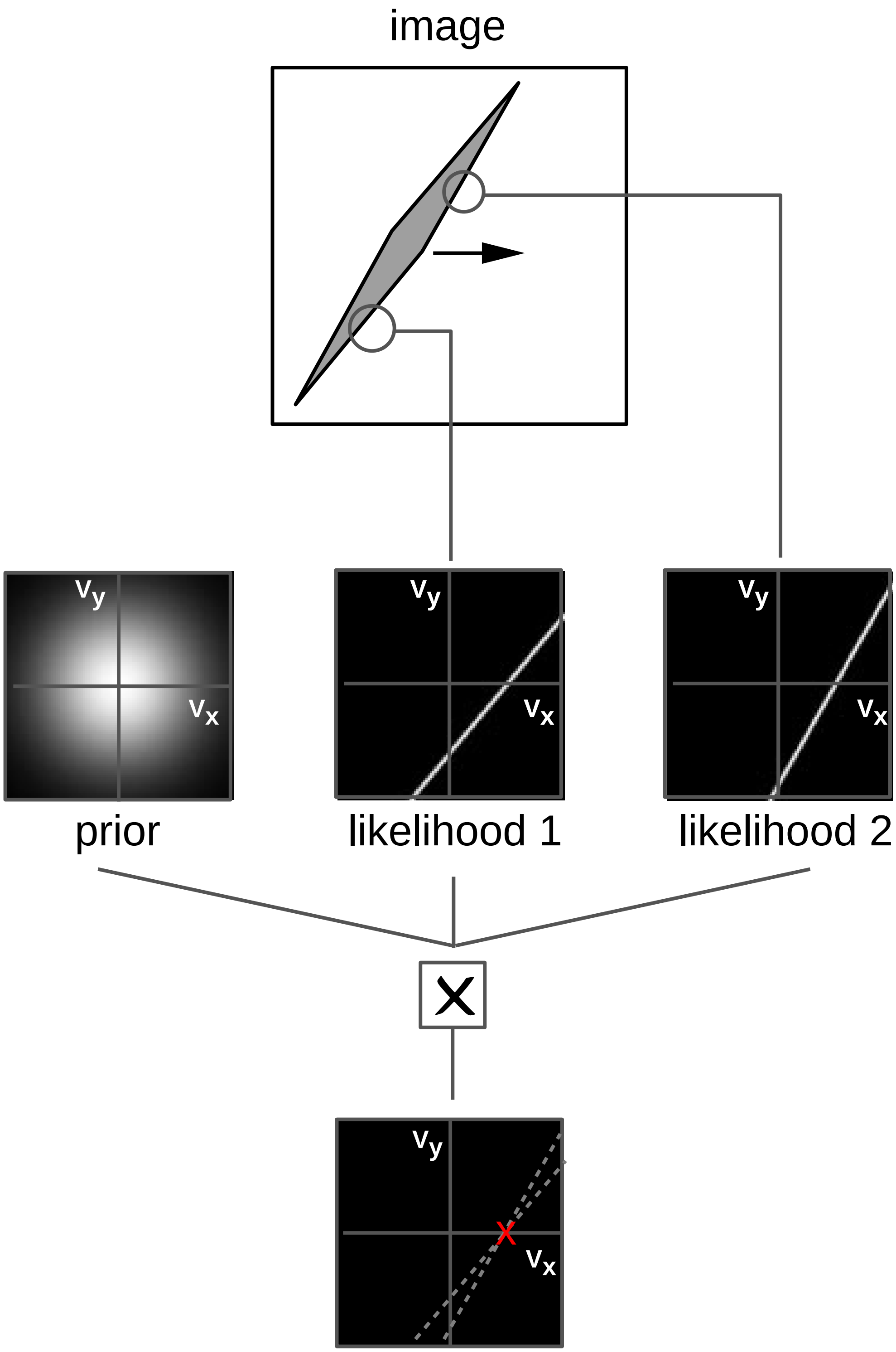


Figure 3: A restricted Bayesian estimator for velocity. The algorithm receives local likelihoods from various image locations and calculates the posterior probability in velocity space. This estimator is too simplistic to account for the range of phenomena we are intersted in explaining but serves to give intuition about how Bayesian motion estimation works. Here the likelihoods are zero everywhere except on the constraint line and the MAP estimate is the IOC solution.

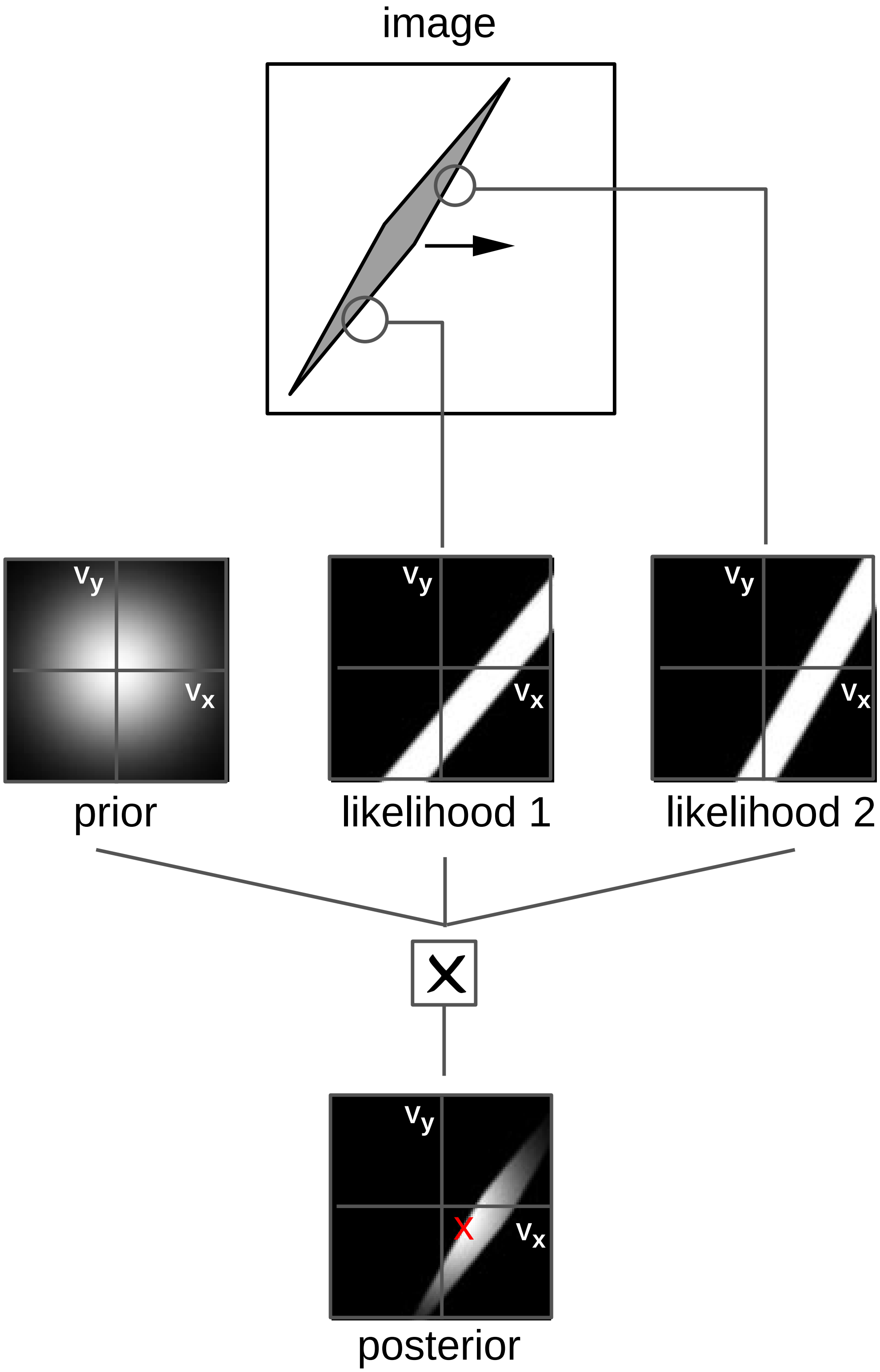


Figure 4: A restricted Bayesian estimator for velocity. The algorithm receives local likelihoods from various image locations and calculates the posterior probability in velocity space. This estimator is too simplistic to account for the range of phenomena we are interested in explaining but serves to give intuition about how Bayesian motion estimation works. Here the likelihoods are zero everywhere except at distance ϵ from the constraint line and the MAP estimate is the normal velocity with minimal speed.

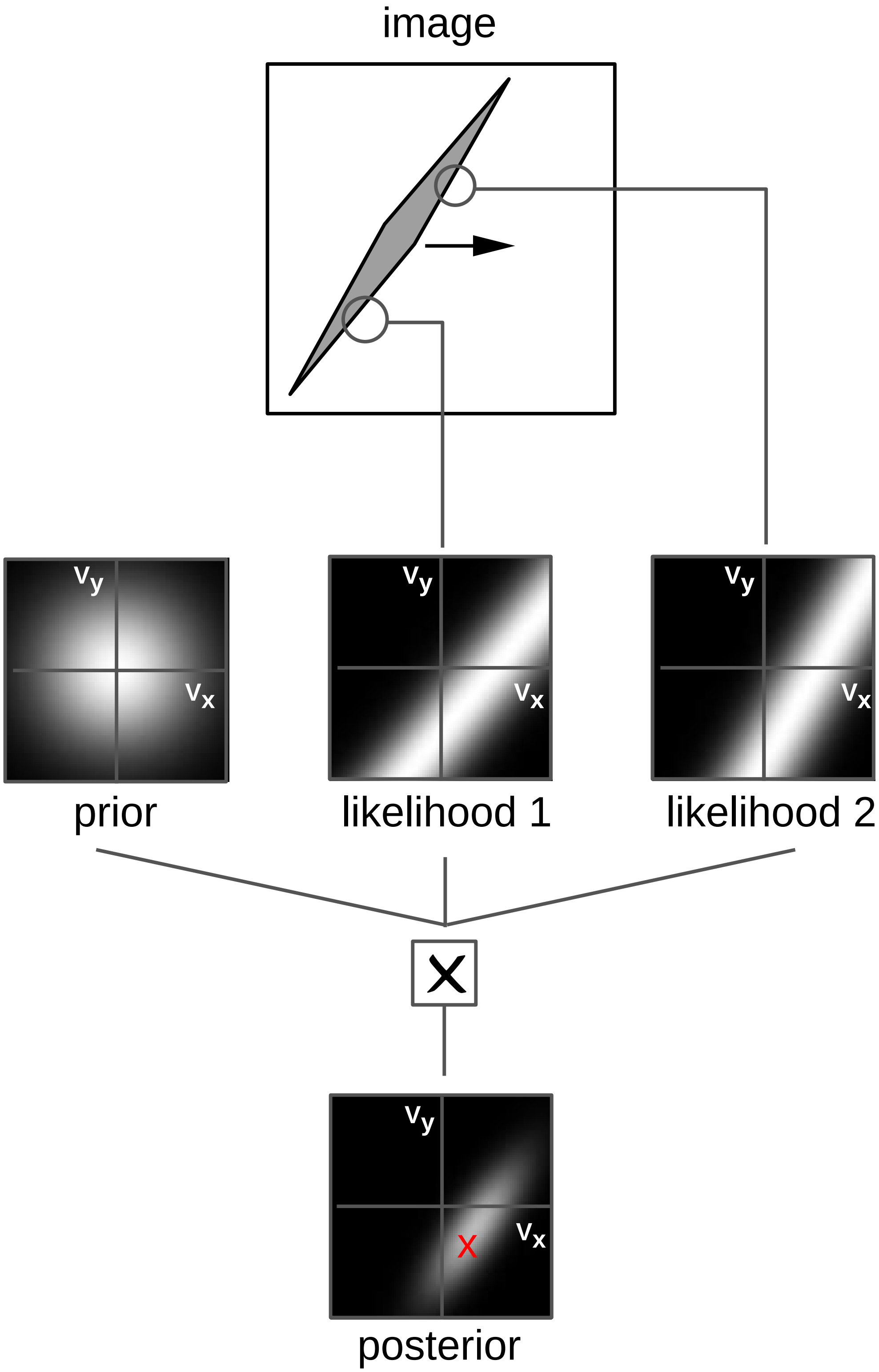


Figure 5: A restricted Bayesian estimator for velocity. The algorithm receives local likelihoods from various image locations and calculates the posterior probability in velocity space. This estimator is too simplistic to account for the range of phenomena we are interested in explaining but serves to give intuition about how Bayesian motion estimation works. Here the likelihoods fall off in a Gaussian manner with distance from the constraint line, and the MAP estimate is the vector average.

order to obtain likelihoods and then multiply these likelihoods and the prior probability to find the posterior.

This suggests the restricted Bayesian motion estimator illustrated in figures 3–5. The model receives as input likelihoods from two apertures, and multiplies them together with a prior probability to obtain a posterior probability in velocity space. Finally the peak of the posterior distribution gives the MAP estimate.

Figure 3 shows the MAP estimate when the two likelihoods are set to 1 for velocities on the constraint line and 0 everywhere else. The prior probability is a Gaussian favoring slow speeds (cf. [11]) — the probability falls off with distance from the origin. In this case, the prior probability plays no role, because when the two likelihoods are multiplied the result is zero everywhere except at the IOC solution. Thus the MAP estimate will be the IOC solution.

A second possibility is shown in figure 4. Here we assume that the likelihoods are zero everywhere except at velocities that are a fixed distance from the constraint line. Now when the two likelihoods are multiplied they give a diamond shaped region of velocity space in which all velocities have equal likelihood. The multiplication with the prior probability gives a "shaded diamond" posterior probability whose peak is shown with a dot. In this case the MAP estimate is the normal velocity of one of the slower grating.

A third possibility is shown in figure 5. Here we assume that the likelihoods are "fuzzy" constraint lines — likelihood decreases exponentially with distance from the constraint line. Now when the two likelihoods are multiplied they give rise to a "fuzzy" ellipsoid in velocity space. The IOC solution maximizes the combined likelihood but all velocities within the "fuzzy" ellipsoid have similar likelihoods. Multiplication with the prior gives a posterior probability whose peak is shown with the X symbol. In this case the MAP estimate is close to the vector average solution.

As the preceding examples show, this restricted Bayesian model may give rise to various velocity space combination rules, depending on the local likelihoods. However, as a model of human perception the restricted Bayesian model suffers from serious shortcomings:

- The likelihood functions are based on constraint lines, i.e. on an experimenter's description of the stimulus. We need a way to calculate likelihoods directly from spatiotemporal data.

- The likelihood functions only consider "1D" locations. We need a way to define likelihoods for all image regions, including "2D" features.

- The velocity space construction of the estimator assumes rigid translation. We need a way of performing Bayesian inference for general motions, including rotations and nonrigid deformations.

In this paper we describe a more elaborate Bayesian estimator. The model works directly on the image data and combines local likelihoods with a prior probability to estimate a velocity field for a given stimulus. The prior probability favors slow and smooth velocity fields. We review a large number of previously published phenomena and find that the Bayesian estimator predicts a wide range of psychophysical results.

3 The model

The global structure of our model is shown in figure 6. As in most motion models, our model can be divided into two main stages - (1) a local measurement stage and (2) a global integration stage where the local measurements are combined to give an estimate of the motion of a surface. For present purposes we also include two stages that are not the focus of this paper - a selection stage and a decision stage.

3.1 Stage 1 - local likelihoods

The local measurement stage uses the output of spatiotemporal filters in order to obtain information about the motion in a small image patch. An important feature of our model is that the filter outputs are not used in order to derive a single local estimate of motion. Rather, the measurements are used to obtain a *local likelihood map* — for any particular candidate velocity we estimate the probability of the spatiotemporal data being generated by that velocity. This stage of our model is very similar to the model proposed by Heeger and Simoncelli (1991) who also suggested a physiological implementation in areas V1 and MT. Here we use a simpler, less physiological version that still captures the important notion of uncertainty in local motion measurements.

There are a number of reasons why different locations have varying degrees of ambiguity. The first reason is geometry. For a location in which the only image data is a straight edge, there are an infinite number of possible velocities that are equally consistent with the local image data (all lying on a constraint line). In a location in which the data is two-dimensional this is no longer the case, and the local data is only consistent with a single velocity.

Thus in the absence of noise, there would be only two types of measurements — "2D" locations which are unambiguous and "1D" locations which have an infinite ambiguity. However when noise is considered all locations will have some degree of ambiguity. In that case one cannot simply distinguish between velocities that "are consistent" with the local image data and those that are not. Rather the system needs to quantify the *degree* to which the data is consistent with a particular velocity.

Here we quantify the degree of consistency using the

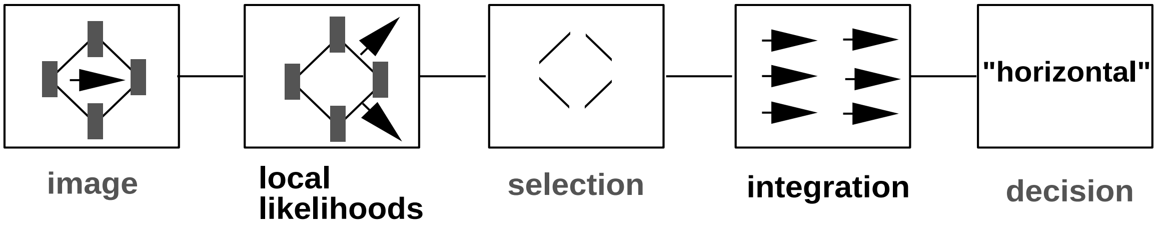


Figure 6: The global structure of our model. Similar to most models of motion perception, our model can be divided into two main stages - (1) a local measurement stage and (2) a global integration stage where the local measurements are combined to give an estimate of object motion. Unlike most models, the first stage extracts *probabilities* about local motion, and the second stage combines these local measurements in a Bayesian framework, taking into account a prior favoring slow and smooth velocity fields.

gradient constraint [13, 16]:

$$C(v_x, v_y) = \sum_{x,y,t} w(x,y,t)(I_x v_x + I_y v_y + I_t)^2 \qquad (5)$$

where v_x, v_y denote the horizontal and vertical components of the local velocity I_x, I_y, I_t denote the spatial and temporal derivatives of the intensity function and $w(x,y,t)$ is a spatiotemporal window centered at (x,y,t). The gradient constraint is closely related to more physiologically plausible methods for motion analysis such as autocorrelation and motion energy [28, 26, 1, 32].

Assuming the intensity of a point is constant as it moves in the image the gradient constraint will be satisfied exactly for the correct velocity. If the local spatiotemporal window contains more than one orientation, the correct velocity can be determined. In the presence of noise, however, the gradient constraint only gives a relative likelihood for every velocity — the closer the constraint is to being satisfied, the more likely that velocity is. A standard derivation under the assumption of Gaussian noise in the temporal derivative [32] gives the likelihood of a velocity at a given location:

$$L(v_x, v_y) = P(I_x, I_y, I_t | v_x, v_y) = \alpha e^{-C(v_x, v_y)/2\sigma^2} \qquad (6)$$

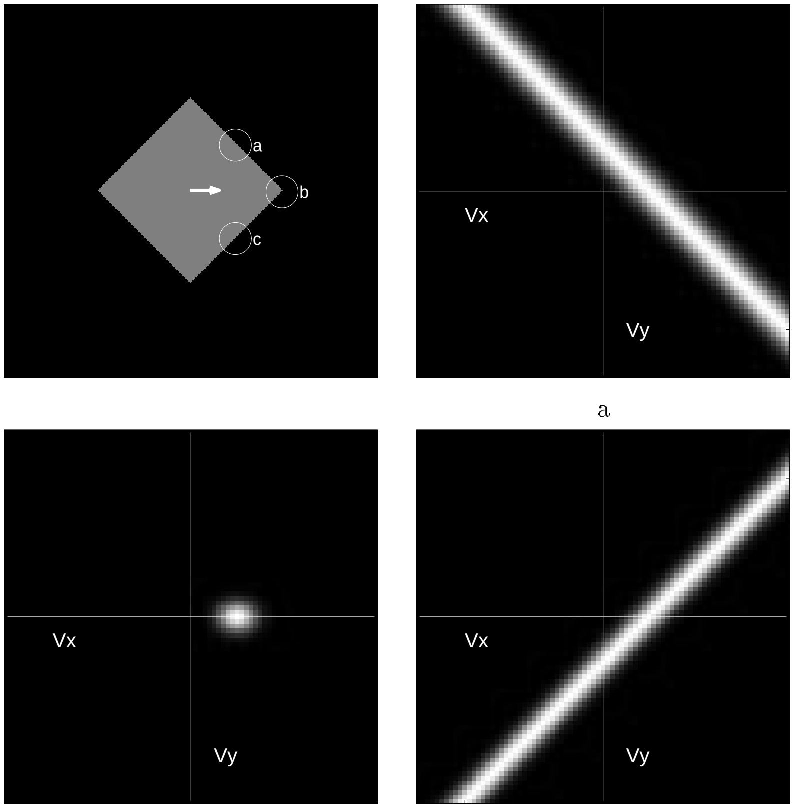
where α is a normalizing constant and σ^2 is the expected variance of the noise in the temporal derivative. This parameter is required in order to convert from the consistency measure to likelihoods. If there is no noise at all in the sequence, then any small deviation from the gradient constraint for a particular velocity means that velocity is extremely unlikely. For larger amounts of noise, the system can tolerate larger deviations from the gradient constraint.

To gain intuition about the local likelihood, we display it as a gray level image for several simple stimuli (figures 7–10). In these plots the brightness at a pixel is proportional to the likelihood of a particular local velocity hypothesis - bright pixels correspond to high likelihoods while dark pixels correspond to low likelihoods.

Figure 7a illustrates the likelihood function at three different receptive fields on a diamond translating horizontally. Note that for locations which have straight lines, the likelihood function is similar to a "fuzzy" constraint line - all velocities on the constraint line have highest likelihood and it decreases with distance from the line. The "fuzziness" of the constraint line is governed by the parameter σ - if we assume no noise in the sequence, $\sigma = 0$, then all velocities off the constraint line have zero, but if we assume noise the falloff is more gradual and points off the constraint line may have nonzero probability. Note also that at corners where the local information is less ambiguous, the likelihood no longer has the elongated shape of a constraint line but rather is centered around the veridical velocity. Our model does not categorize locations into "corners" versus "lines" – all image locations have varying degrees of ambiguity. Figure 8 illustrates the likelihoods at the top of a rotating ellipse. In a "fat" ellipse, the local likelihood at the bottom of the ellipse is highly ambiguous, almost as in a straight line. In a "narrow" ellipse, however, the local likelihood at the bottom of the ellipse is highly unambiguous.

In addition to the local geometry, the uncertainty associated with a location varies with contrast and duration. Although the true velocity will always exactly satisfy the gradient constraint, at low contrasts it will be difficult to distinguish the true velocity from other candidate velocities. The degree of consistency of all velocities will be nearly identical. Indeed in the limiting case of zero contrast, there is no information at all about the local velocity and there is infinite uncertainty. Figure 9 shows the change in the likelihood function *for a fixed* σ as the contrast is varied. At high contrasts the likelihood function is a relatively sharp constraint line, but at lower contrasts it becomes more and more fuzzy — the less contrast the higher the uncertainty. This dependence of uncertainty on contrast is not restricted to the particular choice of consistency measure. Similar plots were obtained using motion energy in [32].

Similarly, the shorter the duration of the stimulus the higher the uncertainty. Since the degree of consistency



a

b

c

Figure 7: A single frame from a sequence in which a diamond translates horizontally. **a-c.** Local likelihoods at three locations. At an edge the local likelihood is a “fuzzy” constraint line, while at corners the local likelihood is peaked around the veridical velocity. In this paper we use the gradient constraint to calculate these local likelihoods but very similar likelihoods were calculated using motion energy in [32]

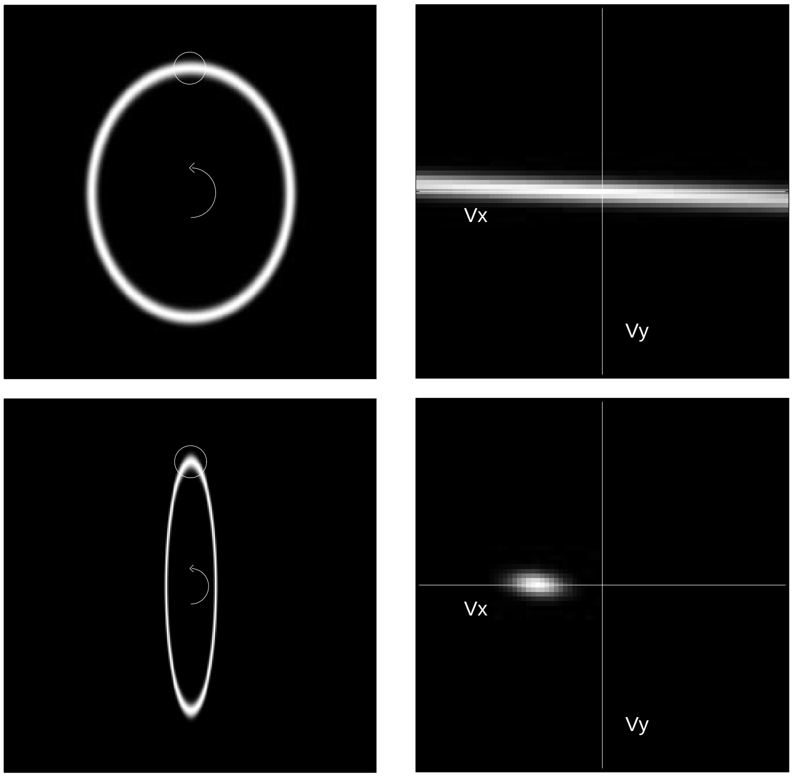


Figure 8: When a curved object rotates, the local information has varying degrees of ambiguity regarding the true motion, depending on the shape. In a “fat” ellipse, the local likelihood at the top of the ellipse is highly ambiguous, almost as in a straight line. In a “narrow” ellipse, however, the local likelihood at the top of the ellipse is relatively unambiguous.

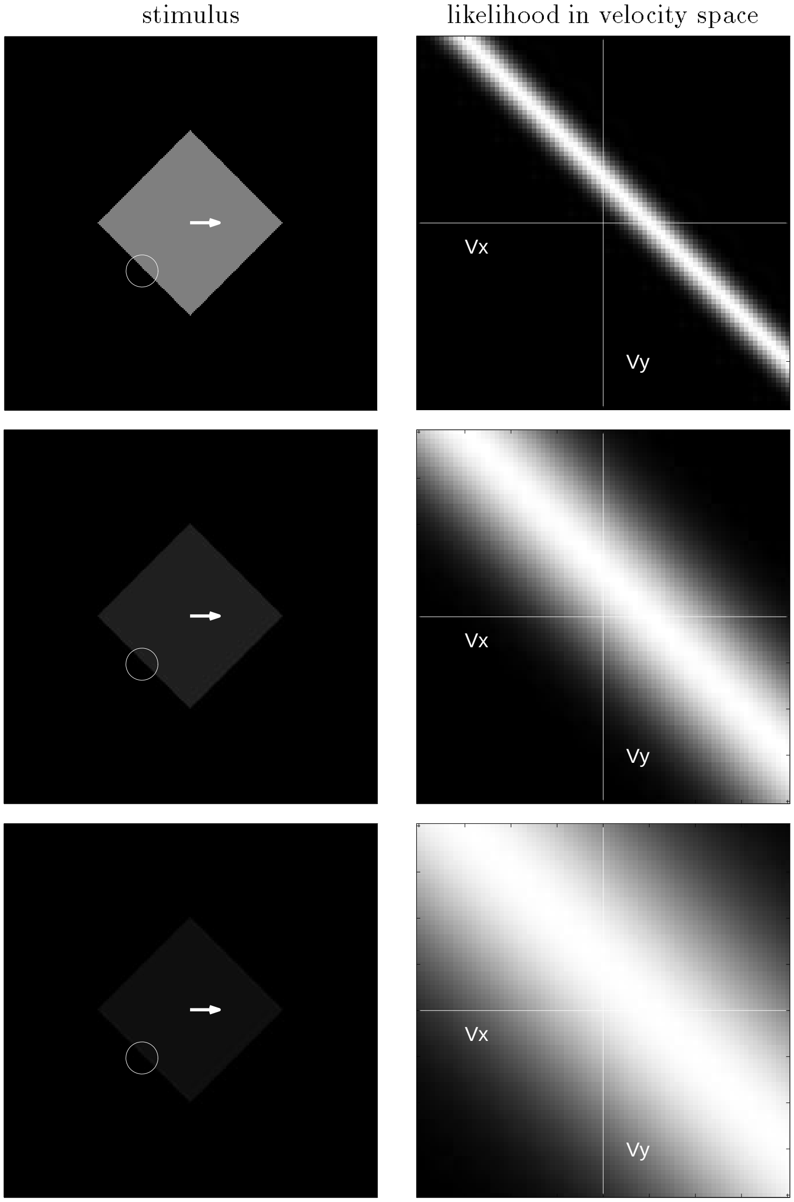


Figure 9: The effect of contrast on the local likelihood. As contrast decreases the likelihood becomes more fuzzy. (cf. [32])
.

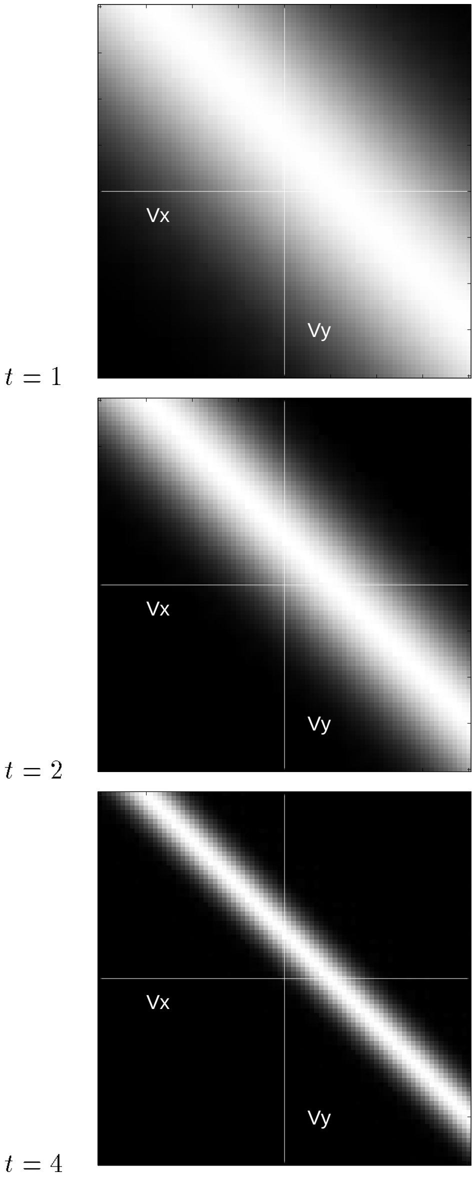


Figure 10: The effect of duration on the local likelihood. As duration increases the likelihood becomes more peaked.

is summed over space and time, it is easier to distinguish the correct velocity from other candidates as the duration of the stimulus increases. Figure 10 illustrates this dependence - as duration increases there is more information in the spatiotemporal receptive field and hence less uncertainty. The likelihood function becomes less fuzzy as duration increases. The quantitative dependence will of course vary with the size and the shape of the window function $w(x, y, t)$, but the

We emphasize again that in the first stage no decision is made about the local velocity. Rather in each local region, a probability distribution summarizes the range of possible velocities consistent with the local data, and the relative likelihood of each of these velocities. The combination of these local likelihoods are left to subsequent processing.

3.2 Stage 2 - Bayesian combination of local signals

Given the local measurements obtained across the image, the second stage calculates the MAP estimate for the motion of a single surface. In the restricted Bayesian model discussed in the introduction, this calculation could be easily performed in velocity space — it required multiplying the likelihoods and the prior to obtain the posterior.

When we consider general motions of a surface, however, the velocity space representation is not sufficient. Any $2D$ translation of a surface can be represented by a single point in velocity space with coordinates (v_x, v_y). However, there is no way to represent a rotation of a surface in a single velocity space plot, we need a larger, higher dimensional space. Figure 11 shows a simple generalization in which motion is represented by three numbers — two translation numbers and a rotation angle. This space is rich enough to capture rotations, but again is not rich enough to capture the range of surface motions — there is no way to capture expansion, shearing or nonrigid deformation. We need a yet higher dimensional space.

We use a 50 dimensional space to represent the motion of a surface. The mapping from parameter space to the velocity field is given by:

$$v_x(x, y) = \sum_{i=1}^{25} \theta_i G(x - x_i, y - y_i) \quad (7)$$

$$v_y(x, y) = \sum_{i=26}^{50} \theta_i G(x - x_i, y - y_i) \quad (8)$$

where $\{x_i, y_i\}$ are 25 locations in the image equally spaced on a $5x5$ grid and $G(x, y)$ is a two dimensional Gaussian function in image space, with spatial extent defined by σ_x:

$$G(x, y) = e^{-\frac{x^2+y^2}{2\sigma_x^2}} \quad (9)$$

There is nothing special about this particular representation — it is merely one choice that allows us to represent a large family of motions with a relatively small number of dimensions. We have also obtained similar results on a subset of the phenomena discussed here with other, less rich, representations.

As in the restricted Bayesian model, we need to define a prior probability over the velocity fields. This is a crucial part of specifying a Bayesian model - after all, one can make a Bayesian model do anything by designing a sufficiently complex prior. Here we choose a simple prior and show how it can account for a wide range of perceptual phenomena.

Our prior incorporates two notions: slowness and smoothness. Suggestions that humans tend to choose the "shortest path" or "slowest" motion consistent with the data date back to the beginning of the century (see [38] and references within). Figure 12a shows two frames of an apparent motion stimulus. Both horizontal and vertical motions are consistent with the information but subjects invariably choose the shortest path motion. Similarly in figure 12b, the wagon wheel may be moving clockwise or counterclockwise but subjects tend to choose the "shortest path" or slower motion. Figure 12c shows an example from continuous motion. The motion of a line whose endpoints are occluded is consistent with an infinite family of velocities, yet subjects tend to prefer the normal velocity, which is the slowest velocity consistent with the data [39].

However, if taken by itself, the bias towards slow speeds would lead to highly nonrigid motion percepts in curved objects. For any image sequence, the slowest velocity field consistent with the image data is one in which each point along a contour moves in the direction of its normal, and hence for objects this would predict nonrigid percepts. A simple example is shown in figure 13 (after Hildreth, 1983). A circle translates horizontally. The slowest velocity field is shown in figure 13b and is highly nonrigid. Hildreth and others [12, 13, 27] have therefore suggested the need for a bias towards "smooth" velocity fields, i.e. ones in which adjacent locations in the image have similar velocities.

To combine the preferences towards (1) slow and (2) smooth motions, we define a prior probability on velocity fields that penalizes for (1) the speed of the velocities and (2) the magnitude of the derivatives of the velocities. Both of these "costs" are summed over the extent of the image. The probability of the velocity field is inversely proportional to the sum of these costs. Thus the most probable velocity field is one in which the surface is static – both the speed and the derivatives of the velocity field are everywhere zero. Velocity fields corresponding to rigid translation in the image plane will also have high probability — since the velocity is constant as a function of space, the derivatives will be everywhere zero. In general, for any candidate velocity field that

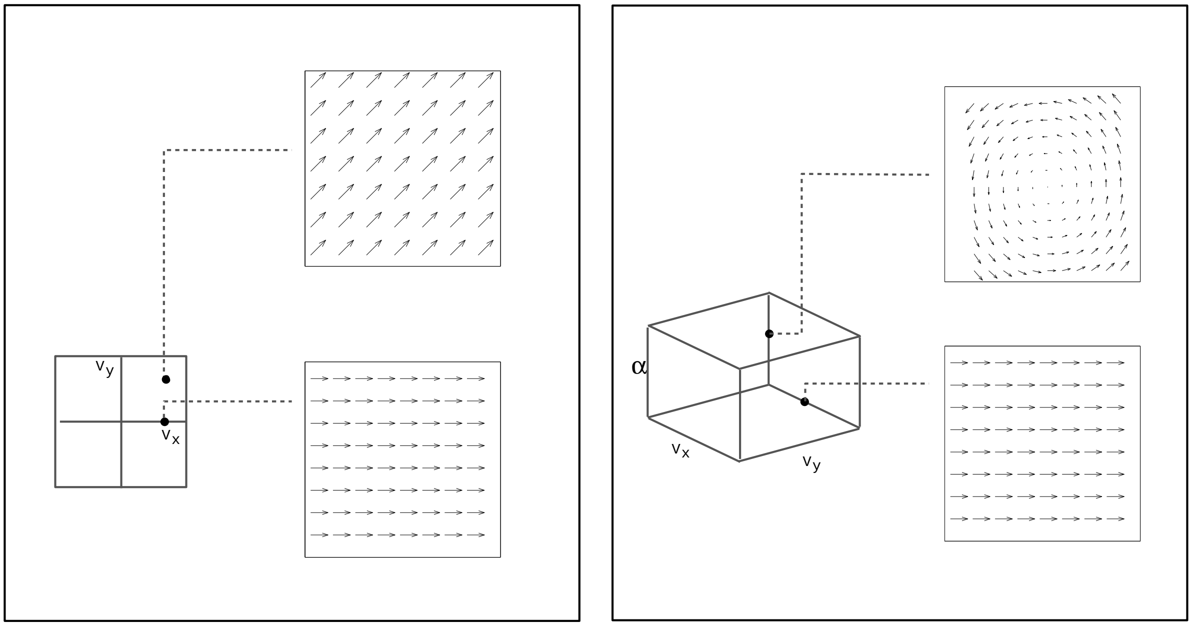


Figure 11: Parametric description of velocity fields. The two dimensional velocity space representation can only represent translational velocity fields. A three dimensional space can represent translational and rotational velocity fields. In this paper we use a 50 dimensional space to represent a rich family of motions including rigid and nonrigid velocity fields.

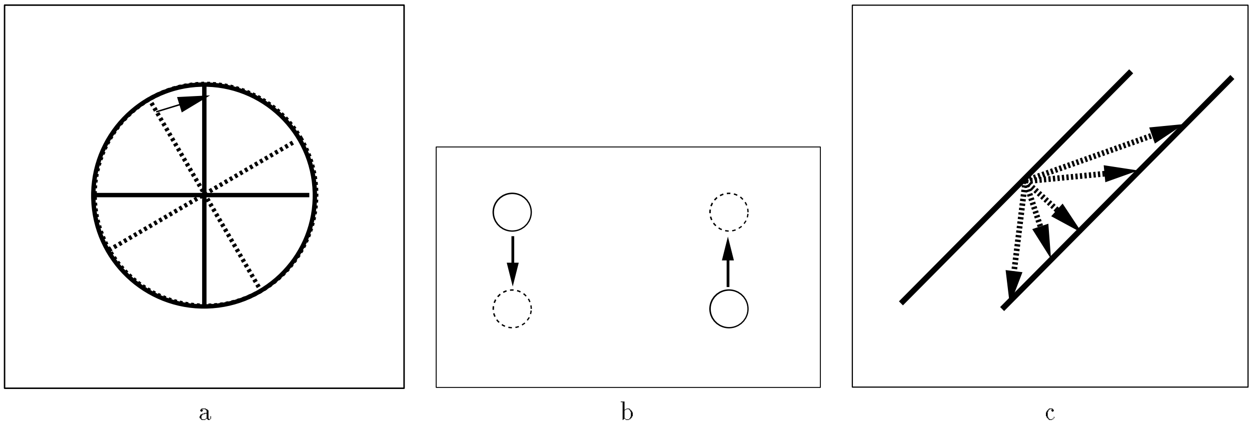


Figure 12: Examples of the preference for slow motions. **a.** A temporally sampled wagonwheel appears to rotate in the shortest direction. **b.** In the “quartet” stimulus, horizontal or vertical motion is perceived depending on which is shortest. **c.** A line whose endpoints are invisible is perceived as moving in the normal, or shortest, velocity.

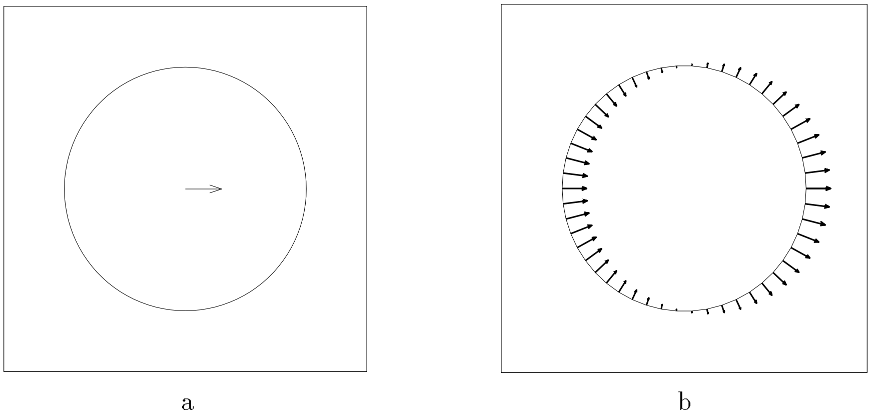


Figure 13: Example of the preference for smooth motions. (after [12]) **a.** A horizontally translating circle. **b.** The slowest velocity field consistent with the stimulus. Based only on the preference towards slower speeds, this stimulus would appear to deform nonrigidly.

can be parameterized by $\vec{\theta}$ we can calculate the prior probability.

Formally, we define the following prior on a velocity field, $V(x, y)$:

$$P(V) = \alpha e^{-J(V)} \tag{10}$$

with:

$$J(V) = \sum_{xy} ||Dv(x, y)||^2 \tag{11}$$

here Dv is a differential operator, i.e. it measures the derivatives of the velocity field. We follow Grzywacz and Yuille (1991) in using a differential operator that penalizes velocity fields with strong derivatives:

$$Dv = \sum_{n=0}^{\infty} a_n \frac{\partial^n}{\partial x} v \tag{12}$$

Note that the sum starts from $n = 0$ thus Dv also includes a penalty for the "zero order" derivative - i.e. it penalizes fast flow fields. For mathematical convenience, Grzywacz and Yuille chose $a_n = \lambda^{2n}/(n!2^n)$ where λ is a free parameter. They noted that similar results are obtained when a_n is set to zero for $n > 2$. We have also found this to be true in our simulations. Thus the main significance of the parameter λ is that it controls the ratio between the penalty for fast velocities ($a_0 = 1$) and the penalty for nonsmooth velocities ($a_1 = \lambda^2/2$). We used a constant value of λ throughout (see appendix).

Unlike the restricted Bayesian model discussed in the introduction, the calculation of the posterior probability cannot be performed graphically. The prior probability of $\vec{\theta}$ for example is a probability distribution over a 50 dimensional space. However, as we show in the appendix it is possible to solve analytically for the most probable $\vec{\theta}$. This gives the velocity field predicted by the model for a given image sequence.

3.3 Selection and Decision

As mentioned in the introduction, in scenes containing multiple objects, the selection of which signals to integrate is a crucial step in motion analysis (cf. [25]). This is not the focus of our paper, but in order to apply our model directly to raw images we needed some rudimentary selection process. We make the simplifying assumption that the image contains a single moving object and (optionally) static occluders. Thus our selection process is based on subtracting subsequent frames and thresholding the subtraction to find regions that are not static. All measurements from these regions are combined. The selection stage also discards all measurements from receptive fields lying exactly on the border of the image, to avoid edge artifacts.

The decision stage is needed in order to relate our model to psychophysical experiments. The motion integration stage calculates a velocity field, but in many experiments the task calls for making a discrete decision based on the perceived velocity field (e.g. "up" versus "down"). In order to model these experiments, the decision stage makes a judgment based on the estimated velocity field. For example, if the experiment calls for a direction of motion judgment, the decision stage fits a single global translation to the velocity field and output the direction of that translation.

3.4 Model Summary

The model starts by obtaining local velocity likelihoods at every image location. These likelihoods are then combined in the second stage to calculate the most probable velocity field, based on a Bayesian prior favoring slow and smooth motions. All results described in the next section were obtained using the Gaussian parameterization (equation 7), with a fixed λ. Stimuli used as input were gray level image sequences (5 frames $128x128$ pixel size) and the spatiotemporal window used to calculate the likelihoods was of size $5x5x5$ pixels.

The only free parameter that varies between experiments is the parameter σ. It corresponds to the observer's assumption about the reliability of his or her temporal derivative estimates. Thus we would expect the numerical value of σ to vary somewhat between observers. Indeed for many of the illusions we model here, individual differences have been reported for the magnitude of the bias (e.g. [45, 15]) although the qualitative nature of the perceptual bias is similar across subjects. Although σ is varied when modeling different experiments, it is always held constant when modeling a single experiment, thus simulating the response of a single observer to varying conditions.

4 Results

We start by showing the results of the model on translating stimuli. Although the Bayesian estimate is a velocity *field*, we summarize the estimate for these stimuli using a single velocity vector. This vector is calculated by taking the weighted mean value of the velocity field with weight decreasing with distance from the center of the image. Except otherwise noted the estimated velocity field is roughly constant as a function of space and is well summarized with a single vector.

4.1 The Barberpole illusion - Wallach 35

Phenomena: As noted by Wallach (1935), a grating viewed through a circular aperture is perceived as moving in the normal direction, but a grating viewed through a rectangular aperture is perceived as moving in the direction of the longer axis of the aperture.

Model Results: Figure 14b,d shows the Bayesian estimate for the two stimuli. In the circular aperture the Bayesian estimate is in the direction of the normal velocity, while in the rectangular one, the estimate is in the direction of the longer axis of the aperture.

Discussion: Recall that the Bayesian estimate combines measurements from different locations according to their uncertainty. For the rectangular aperture, the "terminator" locations corresponding to the edges of the aperture dominate the estimate and the grating is perceived to move horizontally. In the circular aperture, the terminators do not move in a coherent direction, and hence do not have a large influence on the estimate. Among all velocities consistent with the constraint line, the preference for slow speeds favors the normal velocity.

For the rectangular aperture the Bayesian estimate exhibits significant nonrigidity — at the vertical edges of the aperture the field has strong vertical components. We also note that although the present model can account for the basic barberpole effect, it does not account for various manipulations that influence the terminator classification and the magnitude of the barberpole effect. For example, Shimojo et al. (1989) have used stereoscopic depth to place the grating behind the aperture and their subjects tended to perceive the grating as moving closer to the normal direction even in a rectangular aperture. A more sophisticated selection mechanism is required to account for their effect.

4.2 Biases towards VA in translating stimuli

4.2.1 Type II plaids - Yo and Wilson (1992)

Phenomena: Yo and Wilson (1992) distinguished between two types of plaid figures. In "Type I" plaids the two normal velocities lie on different sides of the veridical velocity, while in "type II" plaids both normal velocities lie on the same side and hence the vector average is quite different from the veridical velocity (see figure 15b,d). They found that for short presentation times, or low contrast, the perceived motion of type II is strongly biased in the direction of the vector average while the percept of type I plaids is largely veridical.

Model Results: Figure 15b,d shows the VA, IOC and Bayesian estimate for the two stimuli. For type I plaids the estimated direction is veridical but the speed is slightly slower than the veridical. For type II plaids the Bayesian estimator gives an estimate that is far from the veridical velocity, and that is much closer to the vector average.

Discussion: The decrease speed observed in the Bayesian estimate for type I plaids is to be expected from a prior favoring slow velocities. The bias in direction towards the VA in type II plaids is perhaps less obvious. Where does it come from?

As pointed out by Heeger and Simoncelli (1991), a Bayesian estimate with a prior favoring slow speeds will be biased towards VA in this case, since the VA solution is much slower. Consider figure 15b. Recall that the Bayesian estimate maximizes the product of the likelihood and the prior of the estimate. Let us compare the veridical IOC solution to the Bayesian estimate in these terms.

In terms of **likelihood** the IOC estimate is optimal. It is the only solution that lies exactly on both constraint lines. The Bayesian solution does not maximize the likelihood, since it does not lie exactly on both constraint lines. However, recall that the local likelihoods are "fuzzy" constraint lines, and hence the Bayesian solution which is close to both constraint lines still receives high likelihood. In terms of the **prior**, however, the Bayesian solution is much preferred. It is significantly (about 55%) slower than the IOC solution. Thus a system that maximizes both the prior and the likelihood will not choose the IOC solution, but rather one that is biased towards the vector average.

Note that this argument only holds when the likelihoods are "fuzzy" constraint lines, i.e. when the system assumes some noise in the local measurements. A system that assumed no noise would give zero probability to any velocity that did not lie exactly on both constraint lines and would always choose the IOC solution. Recall that

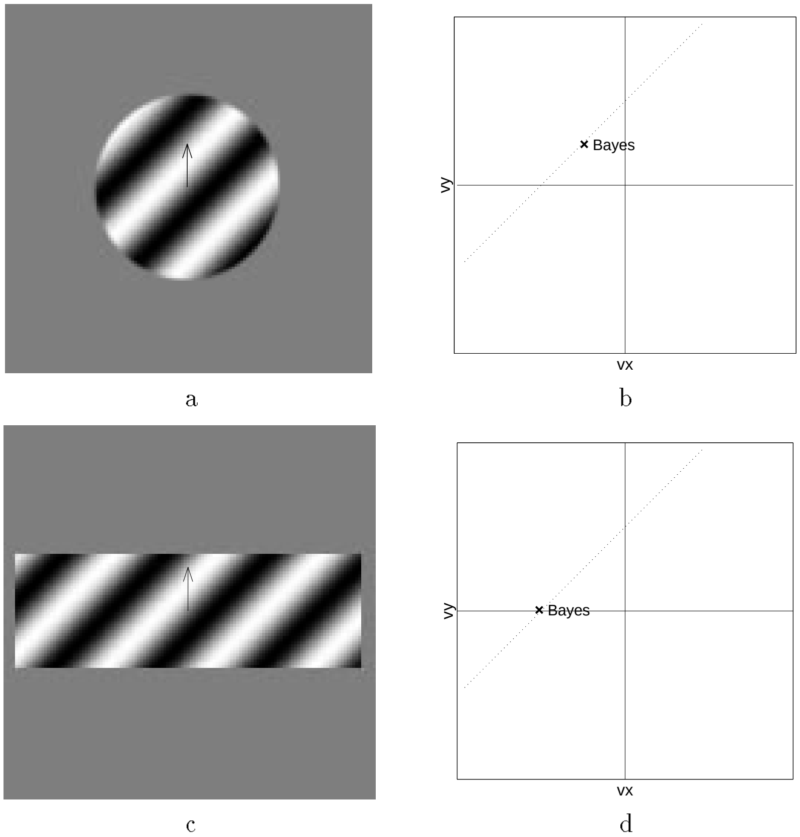


Figure 14: The “barberpole” illusion (Wallach 35). A grating viewed through an (invisible) circular aperture is perceived as moving in the normal direction, but a grating viewed through a rectangular aperture is perceived as moving in the direction of the long axis. **a** A grating viewed through a circular aperture. **b.** The Bayesian estimate for this sequence. Note that the Bayesian estimate is in the normal direction. **c.** A grating viewed through a rectangular aperture. **d.** The Bayesian estimate for this sequence. Note that the Bayesian estimator is now in the direction of the longer axis. Because measurements are combined according to their uncertainty, the unambiguous measurements along the aperture edge overcome the ambiguous ones obtained inside the aperture.

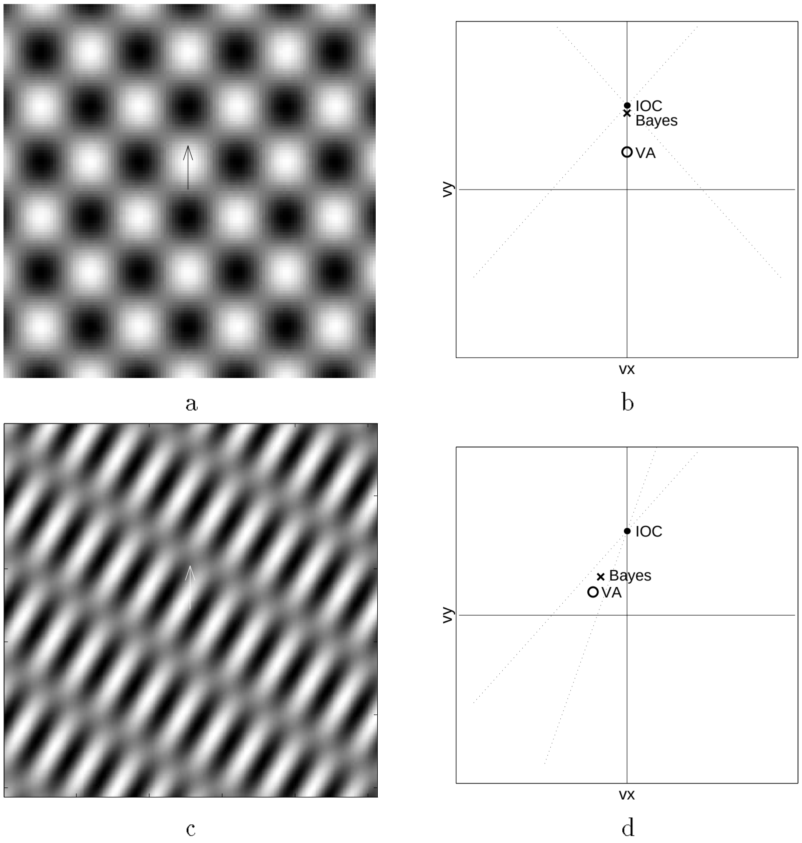


Figure 15: A categorization of plaid patterns introduced by Yo and Wilson (1992). "Type I" plaids have component motions on both sides of the veridical velocity, while "Type II" plaids do not. **a** a "type I" plaid moving upward is typically perceived veridically. **b.** The IOC, VA and Bayesian estimate for this sequence. Note that the Bayesian estimate is in the veridical direction. **c.** a "type II" plaid moving upward is typically perceived to move in the direction of the vector average. **d.** The IOC, VA and Bayesian estimate for this sequence. Note that the Bayesian estimator is biased towards the VA motion, as is the percept of observers. Although the IOC solution maximizes the likelihood, the VA solution has higher prior probability and only slightly lower likelihood.

the degree of "fuzziness" of the constraint lines varies depending on the conditions, e.g. the contrast and duration of the stimulus. Thus the Bayesian estimate may shift from the VA to the IOC solution depending on the conditions. In subsequent sections we show that to be the case.

4.2.2 Biased oriented lines - Mingolla et al. (1992)

Phenomena: Additional evidence for a vector average combination rule was found by Mingolla et al. (1992) using stimuli consisting of lines shown behind apertures (see figure 16a). Behind each aperture, a line translates horizontally, and the orientation of the line is one of two possible orientations. In the "downward biased" condition, the lines are $+15, +45$ degrees from vertical, in the "upward biased" condition, the lines are $-15, -45$ from vertical and in the "no bias" condition the lines are $+15, -15$ degree from vertical. They found that the perceived direction of motion is heavily biased by the orientation of the lines. In a two alternative forced choice experiment, the upward, downward and unbiased line patterns moved in five directions of motion. Subjects were asked to indicate whether the motion was upward or downward. Figures 17a shows the performance of the average subject on this task, replotted from [19]. Note that in the biased conditions, subjects' percept is completely due to the orientation of the lines and is independent of the actual motion.

Model Results: Figure 16b shows the IOC, VA and Bayesian solution for the stimulus shown in figure 16a. The Bayesian solution is indeed biased upwards. Figure 17b shows the 'percent correct' of the Bayesian model in a simulated 2AFC experiment. To determine the percentage of upward responses, the decision module used a "soft" threshold on the velocity field:

$$P = \frac{1}{1 + exp(-\alpha)} \tag{13}$$

where α is the model's estimated direction of motion. This corresponds to a "soft" threshold decision on the model's output. The only free parameter, σ was held constant throughout these simulations. Note that in the biased conditions, the model's percept is completely due to the orientation of the lines and is independent of the actual motion.

Discussion: As in the type II plaid, the veridical velocity is not preferred by the model, due to the prior favoring slower speeds. The veridical velocity maximizes the likelihood but not the posterior. In a second set of simulations (not shown) the terminations of the line endings were visible inside each aperture. Consistent with the results of Mingolla et al. (1992), the estimated direction was primarily a function of the true direction of the pattern and not the orientation.

4.2.3 A manifold of lines (Rubin and Hochstein 92)

Phenomena: Even in stimuli containing more than two orientations, the visual system may be incapable of estimating the veridical velocity. Rubin and Hochstein (1993) presented subjects with a "manifold" of lines translating horizontally (see figure 18a). They asked subjects to adjust a pointer until it matched their perceived velocity and found that the perceived motion was diagonal, in the direction of the vector average. The authors also noted that when a small number of horizontally translating dots were added to the display (figure 18c), the veridical motion was perceived.

Model Results: Figure 18b shows the IOC, VA and Bayesian solution for the manifold stimulus. The Bayesian estimate is biased in the direction of the VA. Figure 18d shows the estimate when a small number of dots are added. The estimate is now veridical.

Discussion: The bias in the absence of features is explained in the previous displays — the veridical velocity maximizes the likelihood but not the posterior. The shift in percept based on a small number of terminator signals falls naturally out of the Bayesian framework. Since individual measurements are combined according to their uncertainty, the small number of measurements from the dots overcome the measurements from the lines.

In Rubin and Hochstein's original displays the lines were viewed through an aperture, unlike the displays used here where the lines fill the image. An interesting facet of Rubin and Hochstein's results which is not captured in our model is that the accidental terminator signals created by the aperture also had a significant effect on the perceived motion. Similar to the results with the barber pole illusion, they found that manipulating the perceived depth of the aperture changed the influence of the terminators. A more sophisticated selection mechanism is needed to account for these results.

4.2.4 Intermediate solutions - Bowns (1996)

Phenomena: The Bayesian estimator generally gives a velocity estimate somewhere between "pure" vector average and "pure" IOC. Evidence against either pure mechanism was recently reported by Bowns (1996). In her experiment, a set of type II plaids consisting of orientations 202 and 225 were used as stimuli. Although the two orientations were held constant, the relative speeds of the two components were varied. The result was a set of plaids where the vector average was always right of the vertical while the IOC solution was always left of vertical. Figure 19 shows examples of the two extreme plaids used in her study, along with their velocity space construction.

Subjects were asked to determine whether or not the motion was left or right of vertical. It was found that when the speeds of the two components were similar, subjects answered right of vertical (consistent with the

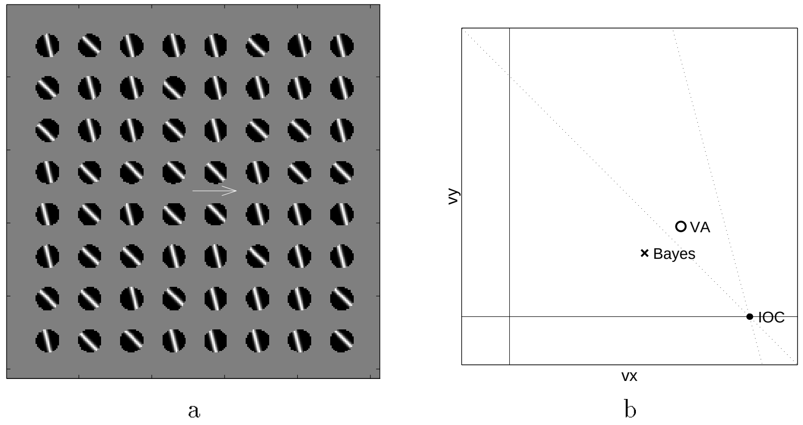


Figure 16: A stimulus studied by Mingolla et al. (1992) suggesting a vector average combination rule. **a.** a single frame from a sequence in which oriented lines move horizontally behind apertures. **b.** The IOC, VA and Bayesian estimate for this sequence. Note that the Bayesian estimator is biased towards the VA motion, as is the percept of observers [19].

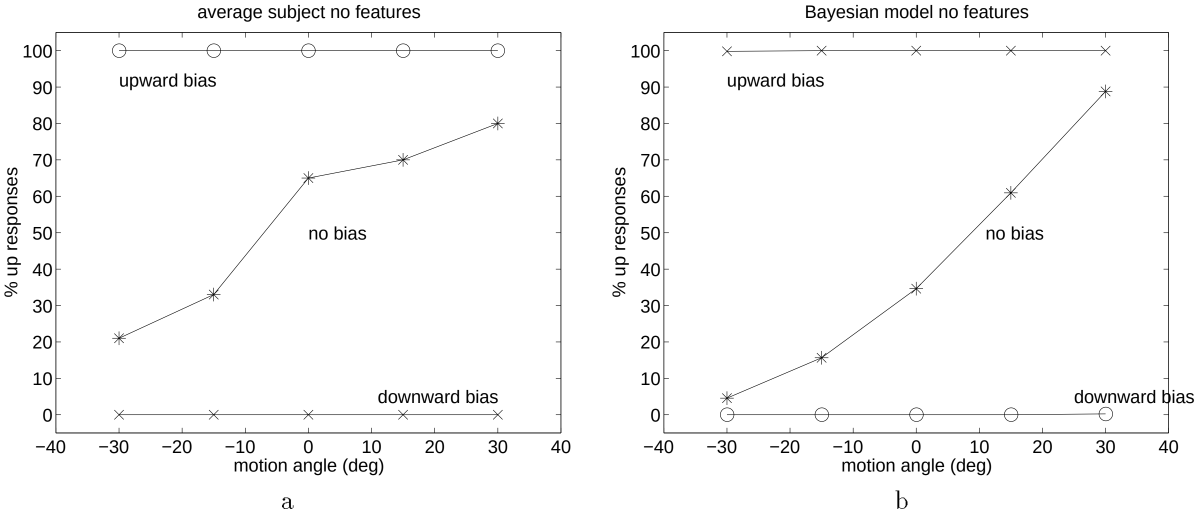


Figure 17: **a.** Results of experiment 1 in [19]. Three variations on the line images shown in figure 16a moved in five directions of motion. Subjects were asked to indicate whether the lines moved upward or downward. Note that in the absence of features, the perceived direction was only a function of the orientation of the lines. **b.** Results of Bayesian estimator output on the same stimuli. The single free parameter σ is held constant throughout.

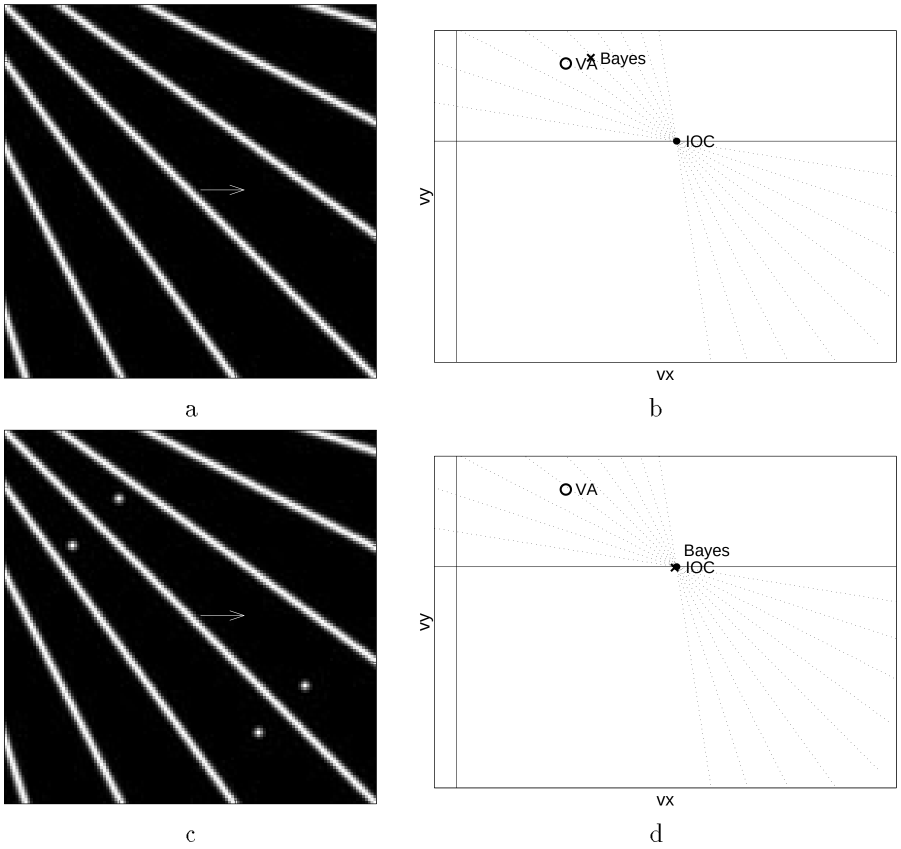


Figure 18: **a.** A single frame from a stimulus introduced by Rubin and Hochstein (1993). A collection of oriented lines translate horizontally. **b.** The VA, IOC and Bayesian estimate. The Bayesian estimate is biased in the vector average direction, consistent with the percept of human subjects. **c.** When a small number of dots are added to the display the pattern appears to translate horizontally (Rubin and Hochstein 92). **d.** The Bayesian estimate shifts to veridical under these circumstances. Since individual measurements are combined according to their uncertainty, the small number of measurements from the dots overcome the measurements from the lines.

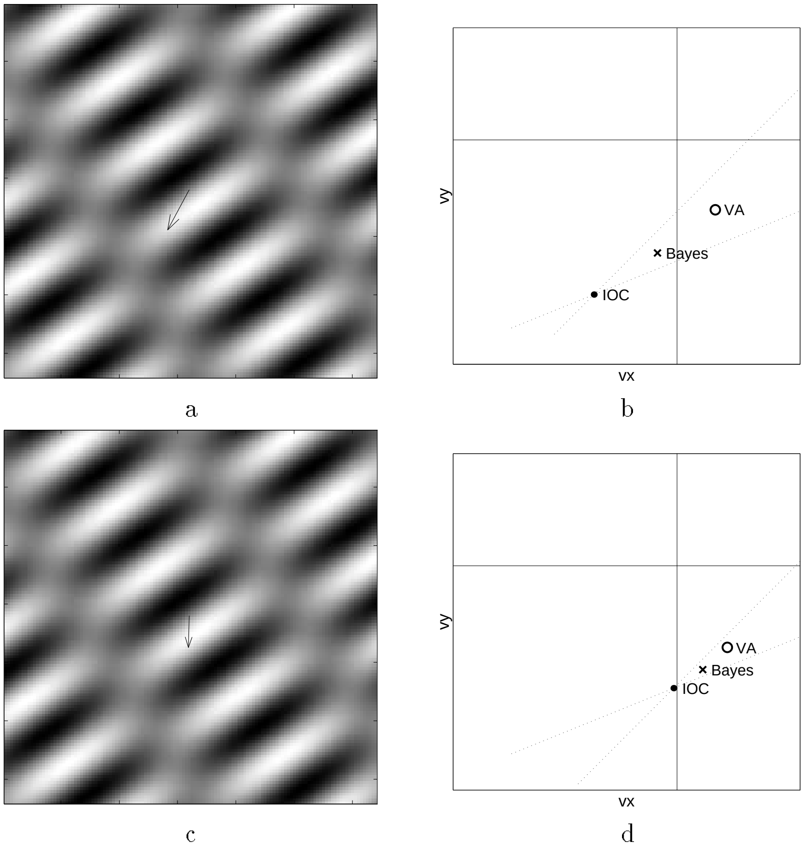


Figure 19: Experimental stimuli used by Bowns (1996) that provide evidence against a pure vector average or IOC mechanism. **a.** A type II plaid with orientations $202, 225$ degrees and relative speeds $1, 0.45$. **b.** The VA, IOC and Bayesian estimates. The IOC solution is leftward of the vertical while the VA solution is rightward. The Bayesian estimate is leftward, consistent with the results of Bowns (1996). **c.** A type II plaid with orientations $202, 225$ degrees and relative speeds $1, 0.75$. **d.** The VA, IOC and Bayesian estimates. The IOC solution is leftward of the vertical while the VA solution is rightward. The Bayesian estimate is rightward, consistent with the results of Bowns (1996).

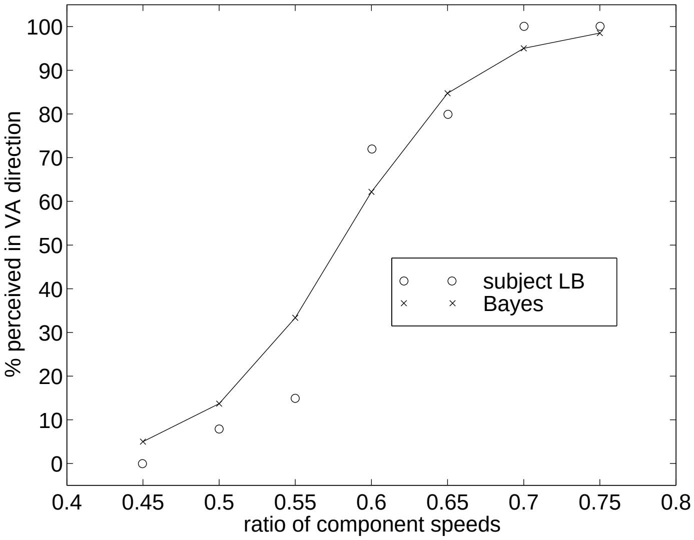


Figure 20: The results of an experiment conducted by Bowns (1996). Subjects indicated if the motion of a plaid was left of vertical (consistent with VA) or rightwards of vertical (consistent with IOC). The relative speeds of the two components were varied. The circles show the results of subject LB, while the crosses show the output of the Bayesian model (σ constant throughout). The experimental results are inconsistent with pure VA or pure IOC but are consistent with a Bayesian estimator.

VA solution) while when the speeds were dissimilar subjects answered left of vertical (consistent with the VA solution). The circles in figure 20 show the percentage of rightward results for a subject in her experiment.

Model Results Figure 19c and d show the Bayesian estimate for the two extreme cases. Note that they switch from left of vertical to right of vertical as the relative speeds change. In figure 20 the solid line gives the expected percent rightward responses for the Bayesian estimator. Note that it gives a gradual shift from left to right as the relative speeds are varied. The parameter σ is held constant throughout.

Discussion: Here again, the prior favoring slower speeds causes the Bayesian estimator to move away from the veridical IOC solution. However, the Bayesian estimator is neither a "pure" IOC solution nor a "pure" VA solution. Rather it may give any perceived velocity that varies smoothly with stimulus parameters.

The fact that a Bayesian estimator is biased towards the vector average solution suggests that the VA bias is not a result of the inability of the visual system to correctly solve for the IOC solution, but rather may be a result of a combination rule that takes into account noise and prior probabilities to arrive at an estimate.

4.3 Dependence of VA bias on stimulus orientation

4.3.1 Effect of component orientation - Burke and Wenderoth (1992)

Phenomena: Even in type II plaids, the perceived direction may be more consistent with IOC than VA [4, 7]. Consider, for example, the type II plaids shown in figure 21. Burke and Wenderoth (1993) found that for the plaid in figure 21a (orientations $200, 210$) the perceived direction is biased by about 15 degrees, while for the plaid in figure 21c (orientations $185, 225$) the perceived direction is nearly veridical with a bias of under 2 degrees. Thus if one assumes independent IOC and VA mechanisms, one would need to assume that the visual system uses the IOC mechanism for the plaid in figure 21c but switches to the VA mechanism for the plaid in figure 21a. Burke and Wenderoth systematically varied the angle between the two plaid components and asked subjects to report their perceived directions. The results are shown in the open circles in figure 22. The perceived direction is inconsistent with a pure VA mechanism or a pure IOC mechanism. Rather it shows a gradual shift from the VA to the IOC solution as the angle between the components increases.

Model Results: Figure 22 shows the predicted IOC, VA and Bayesian estimates as the angles are varied. The parameter σ is held fixed. Note that a single model generates the range of percepts, consistent with human observers.

Discussion: To get an intuitive understanding of why the same Bayesian estimator gives IOC or VA type solutions depending on the orientation of the components, compare figure 21b to figure 21d. Note that in figure 21b the two constraint lines are nearly parallel. Hence, a solution lying halfway between the two constraint lines

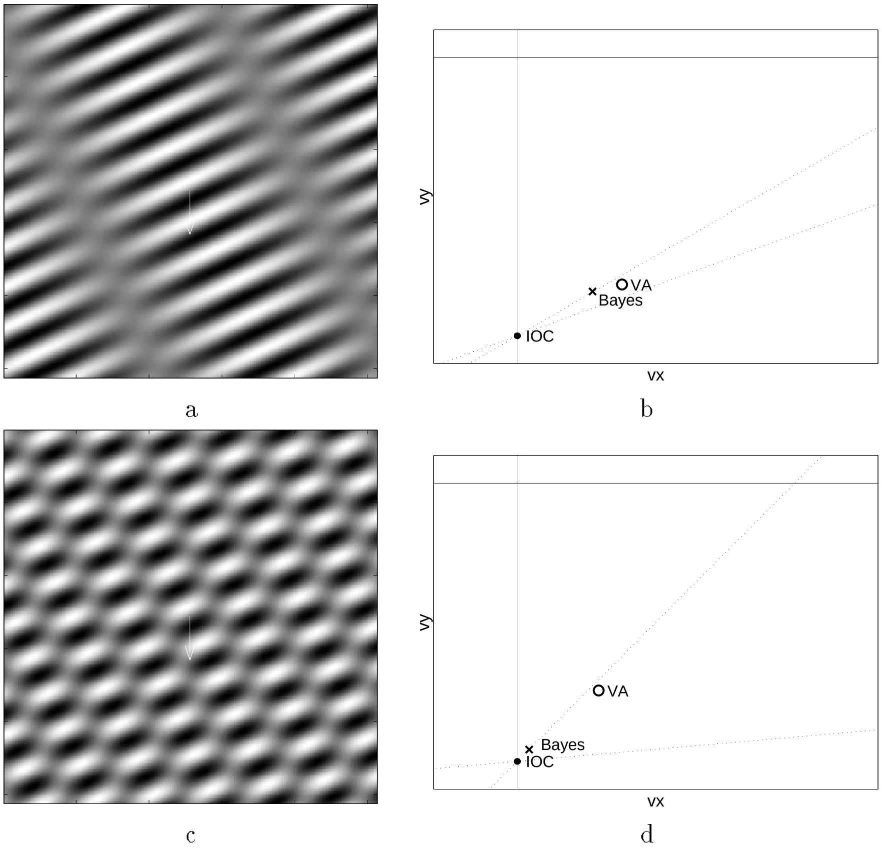


Figure 21: Stimuli used by Burke and Wenderoth (1993) to show that the percept of some type II plaids is more consistent with IOC than with VA. **a.** A type II plaid with orientations 20 and 30 degrees is misperceived by about 15 degrees.[7] **b.** The VA, IOC and Bayesian estimates. The Bayesian estimate is biased in a similar manner to the human observers. **c.** A type II plaid with orientations 5 and 45 degrees is is perceived nearly veridically. [7] **d.** The VA, IOC and Bayesian estimate. The Bayesian estimate is nearly veridical. The parameter σ is held constant throughout.

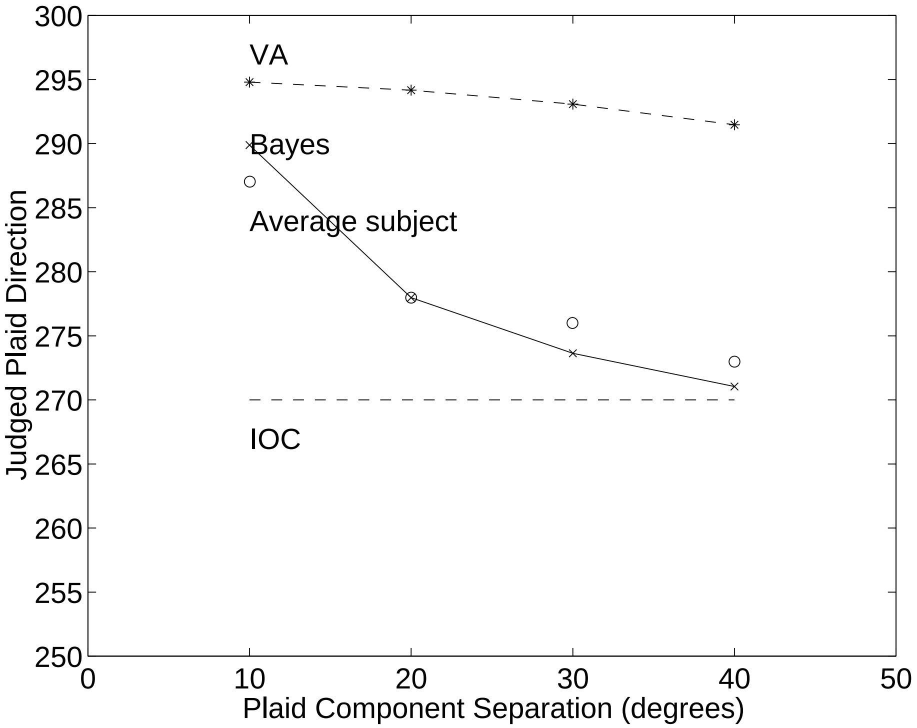


Figure 22: Results of an experiment conducted by Burke and Wenderoth (1993) to systematically investigate the effect of plaid component orientation on perceived direction. All they plaids are "type II" and yet when the relative angle between the components of the plaid is increased varied, the perceived direction shows a gradual shift from the VA to the IOC solution (open circles replotted from [7]). The Bayesian estimator, with a fixed σ shows the same behavior

(such as the VA solution) receives high likelihood for fuzzy constraint lines. However, in figure 21d, where the components have a 40 degrees difference in orientation, the two constraint lines are also separated by 40 degrees. Thus for a solution to have high likelihood, it is forced to lie close to the intersection of the two lines, or the IOC solution. Thus a single, Bayesian mechanism predicts a gradual shift from VA to IOC as the orientation of the components is varied.

An alternative explanation of the shift in perceived direction was suggested by Bowns (1996) who pointed out that there exist features in the "blob" regions of these plaids that move in different directions as the orientations of the gratings are varied. Our results do not of course rule out this explanation, but they show that hypothesizing a specialized "blob" mechanism is not necessary.

4.3.2 Orientation effects in occluded stimuli

Phenomena: We have performed experiments with the stimulus shown in figure 23a. A rhombus whose four corners are occluded is moving horizontally. Note that there are no features on this stimulus which move horizontally - the two normal velocities are diagonal and the terminator motion is downward. This stimulus is similar to a type II plaid in the sense that the two normal velocities lie on the same side of the veridical velocity. However it requires integration across space rather than across multiple orientations at a single point. We wanted to see whether the biases in perceived velocity would behave the same way as in plaids.

We presented subjects with these stimuli while varying the angle of one of the sides and asked them to indicate the perceived direction. Results of a typical subject are shown in figure 23. Consistent with the result on plaids [7] subjects percept shift gradually from a bias in the VA direction to the veridical direction as the angular difference increases.

Model Results: Figure 23 shows the result of the Bayesian estimator with fixed σ. Similar to the results with plaids, the Bayesian estimate shifts gradually from a bias in the VA direction to the veridical direction as the angular difference increases.

Discussion: It seems difficult to reconcile these results with a "multiple mechanism" model in which the visual system uses a VA mechanism or an IOC mechanism depending on the conditions. First, one would have to assume that the visual system uses a different mechanism for nearly identical stimuli, when the relative orientations is changed. Second, the perceived direction changes continuously and includes intermediate values that are inconsistent with either VA or IOC.

Likewise, these results are difficult to reconcile with a "feature tracking" explanation of the sort proposed by Bownes (1996) or by Yo and Wilson (1992) . No matter what the orientation of the rhombus sides are, there are never any trackable features moving in the veridical direction. Yet subjects perceive motion in the veridical direction when the angle between the two components is large.

In contrast, as we have shown, these results are consistent with a Bayesian estimation strategy where motion signals are fused in accordance with their uncertainty and combined with a prior favoring slow and smooth velocities. Again, this does not rule out the "multiple mechanism" explanation, but shows that it is not necessary. A single Bayesian mechanism is sufficient.

4.4 Dependence of VA bias on contrast

4.4.1 Effect of contrast on type II plaids - Yo and Wilson (1992)

Phenomena: Yo and Wilson (1992) reported that the bias towards VA in type II plaids consistently increased with reduced contrast. For example, Figure 24a,c show a type II plaid at high contrast and at low contrast. For durations over 100msec the high contrast plaid is perceived as moving in the veridical direction, while the low contrast is heavily biased towards the VA solution [45].

Model Results: Figure 24b,d show the VA, IOC and Bayesian predictions for this stimulus. Obviously, both VA and IOC solutions are unaffected by the contrast and hence cannot by themselves account for the percept. The Bayesian estimate, on the other hand, changes from veridical to biased as contrast is decreased even though the only free parameter σ is held constant.

Discussion: To gain intuitive understanding of the change in the Bayesian prediction as contrast as varied, recall from section 3.1 that the contrast changes the "fuzziness" of the constraint line. Thus at low contrast, both constraint lines are very fuzzy, and the VA solution receives relatively high likelihood relative to the IOC solution. We emphasize that this change in "fuzziness" with contrast does not have to be put in especially to explain this phenomena. It is a direct consequence of the probabilistic formulation – at low contrast there is more uncertainty locally. Figure 25a shows the consistency measure (equation 5) for different vertical velocities measured at a single location in the stimulus shown in figure 24. At low contrast there is only a small difference between the degree to which the true velocity satisfies the gradient constraint and the degree to which other velocities do so. Therefore when the local likelihoods are calculated (equation 6) one obtains figure 25b. At lower contrast the likelihood function is less peaked, and there is more local uncertainty.

While the sharpness of the local likelihoods change with contrast, the prior probability does not change. As mentioned earlier, the prior probability of the VA solution is higher, and hence at low contrasts the Bayesian solution is biased towards the VA. At high contrast, however, as the likelihoods become much more peaked, the prior has less influence and the Bayesian estimate ap-

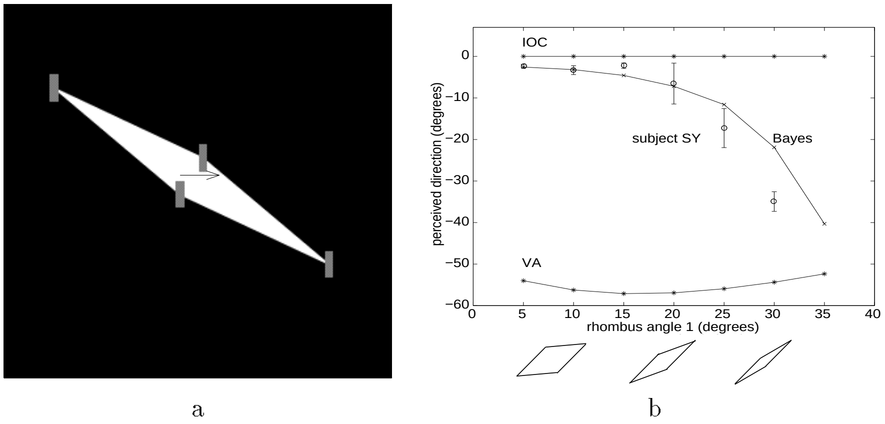


Figure 23: The stimulus used in an experiment to measure influence of relative orientation on perceived direction. A rhombus whose four corners are occluded was translating horizontally. The angle between the two orientations was varied. **a.** A single frame from the sequence. **b.** The predictions of VA, IOC and the Bayesian estimator for the direction of motion of the rhombus. One of the orientations is fixed at 40 degrees, and the second orientation is varied. The VA solution is always far from horizontal (by at least 50 degrees), the IOC prediction is always horizontal and the Bayesian estimator predicts a gradual shift from horizontal to diagonal as the angle between the two components is decreased. The results of a single subject are shown in circles.

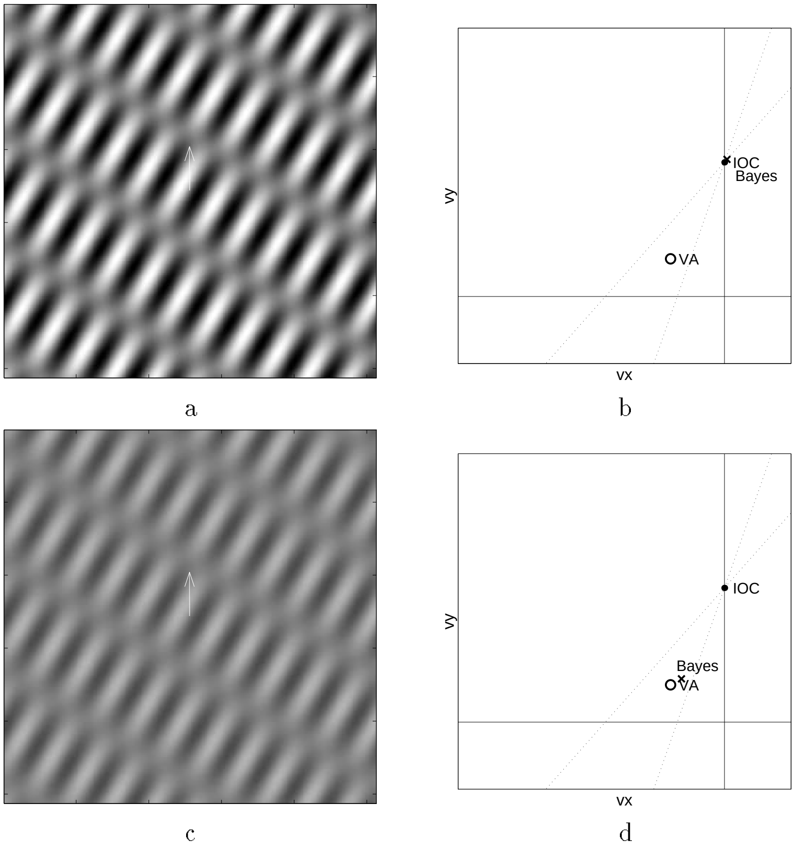


Figure 24: A high contrast type II plaid (a) viewed at long durations, may be perceived veridically, but the same stimulus at low contrast (b) shows a strong VA bias (Yo and Wilson 92). As shown in (b) and (d) the VA and IOC predictions are not affected by contrast, but the Bayesian estimator with a fixed σ shows the same shift from veridical to biased as contrast is decreased.

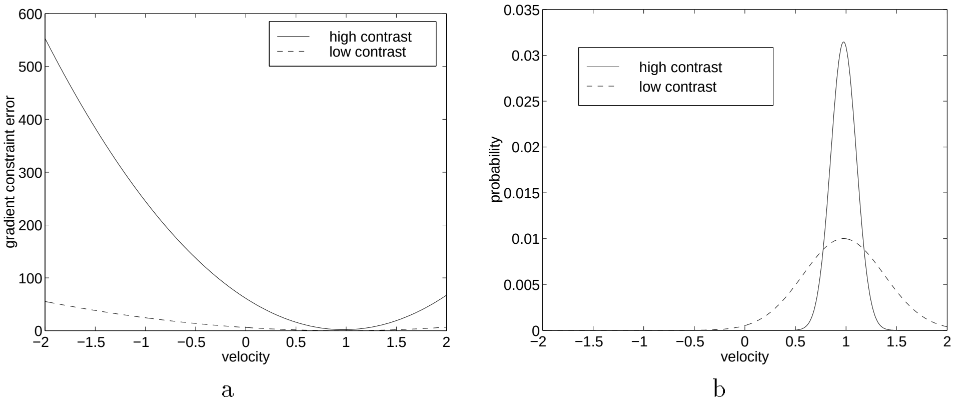


Figure 25: **a.** The local consistency (equation 5) for various vertical velocities measured at a single location in the stimulus shown in figure 24. At low contrast there is only a small difference between the degree to which the true velocity satisfies the gradient constraint and the degree to which other velocities do so. **b.** The local likelihood (equation 6) for various vertical velocities at the same location. At low contrast there is a higher degree of uncertainty.

proaches the IOC solution.

4.5 Contrast effects on line stimuli - Lorenceau et al 1992

Phenomena: Lorenceau et al. (1993) asked subjects to judge whether a matrix of oriented lines moved above or below the horizontal (see figure 26a) as the contrast of the display was systematically varied. The results are replotted in figure 26b. Note that at low contrasts, performance is far below chance indicating subjects perceived upward motion while the patterns moved downward. Lorenceau et al. modeled these results using two separate mechanisms, one dealing with terminator and other with line motion. The terminator mechanism is assumed to be active primarily at high contrast and the line mechanism at low contrast.

Model Results: The solid line in figure 26b shows the simulated performance of the Bayesian model on this task. Again, the percentage of correct responses is obtained by using a "soft" threshold on the model's predicted direction of motion. Although the model does not include separate "terminator" and "line" motion mechanisms, it predicts a gradual shift from downward motion to upward motion as contrast is increased. The parameter σ is held fixed.

Discussion: The intuition behind the model's performance in this task is similar to the one in the plaid displays. At high contrast, the likelihood is peaked and the estimated motion is veridical. At low contrast, however, the likelihood at the endpoints of the lines and along the lines, is more "fuzzy" and the prior favoring slow velocities has a large influence. Hence, motion is perceived in the normal velocity which is slower than the veridical one. There is no need to assume separate terminator and line mechanisms.

4.5.1 Influence of contrast on the speed of a single grating - Thompson et al 1996

Phenomena: Thompson et al. (1996) have shown that the perceived speed of a single grating depends on the contrast. Noting that "lower-contrast patterns consistently appear to move slower", they conducted an experiment in which subjects viewed a high contrast (70%) grating followed by a test low contrast (10%) grating. The subjects adjusted the speed of the test grating until the perceived speeds were matched (see figure 27a). Although the magnitude of the effect varied slightly between subjects, the direction of the effect was quite robust. Typical results are shown in figure 27b. In order to match the perceived speed of the low contrast grating, the high contrast grating needs to move about 70% slower. Similarly, in order to match the perceived speed of the high contrast grating, the low contrast grating needs to move about 150% faster.

Model Results: Figure 27c shows the output of a Bayesian estimator on this stimulus. For a fixed σ the low contrast grating is predicted to move slower. The predicted speed match is computed by dividing the estimated speeds of the two gratings.

Discussion: Again, at at low contrast the likelihood is less peaked and the prior favoring slow speeds dominates. Hence the low contrast grating is predicted to move slower than a high contrast grating moving at the same speed.

4.5.2 Dependence of type I direction on relative contrast - Stone et al. (1990)

Phenomena: Stone et al. (1990) showed subjects a set of type I plaids and varied the ratio of the contrasts between the two components. They found that the direction of motion of the plaid was biased in the direction of the higher contrast grating. The magnitude of the bias changed as a function of the "total contrast" of the

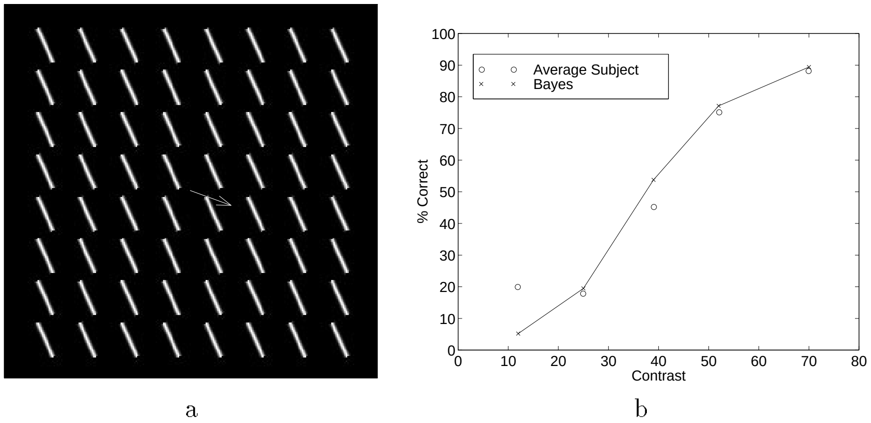


Figure 26: A stimulus used by Lorenceau et al. (93) suggesting the need for independent terminator and line motion mechanisms. A matrix of lines moves oblique to the line orientations. At high contrast the motion of the lines is veridical while at low contrast it is misperceived **a.** A single frame from the sequence. **b.** The results of a two alternative forced choice experiment (up/down) replotted from Lorenceau et al. (1992) (average subject shown with circles). The solid line shows the predictions of the Bayesian model. A single Bayesian mechanism would predict systematic errors at low contrast with an increase in correct responses as contrast is increased.

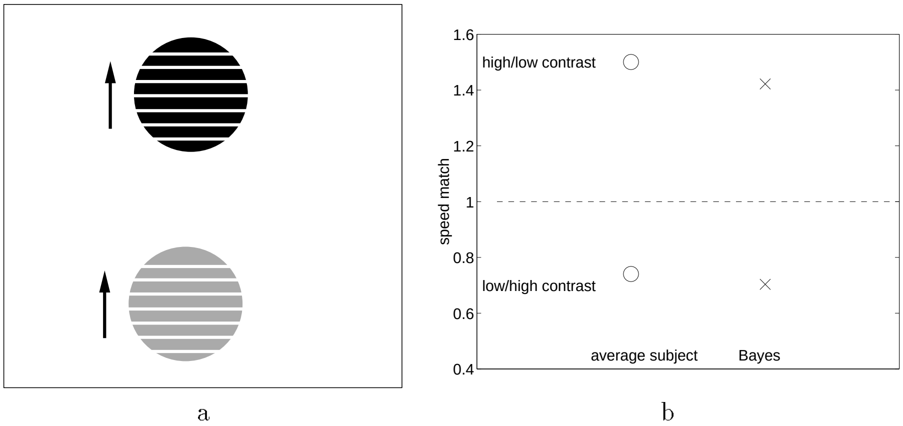


Figure 27: An experiment conducted by Thompson et al. (1996) showing that low contrast stimuli appear to move slower. Subjects viewed a high contrast grating (70%) followed by a test low contrast grating (10%). They adjusted the speed of the test grating until the perceived speeds were matched. **b.** Circles show the results averaged over 6 subjects replotted from [36]. In order to match the perceived speed of a low contrast grating, the high contrast grating needs to move about 70% slower. Similarly, in order to match the perceived speed of a high contrast grating, the low contrast grating needs to move about 150% faster. Crosses show the output of the Bayesian estimator. At low contrast, the likelihood is less peaked and the prior favoring slow speeds dominates. Hence the low contrast grating is predicted to move slower.

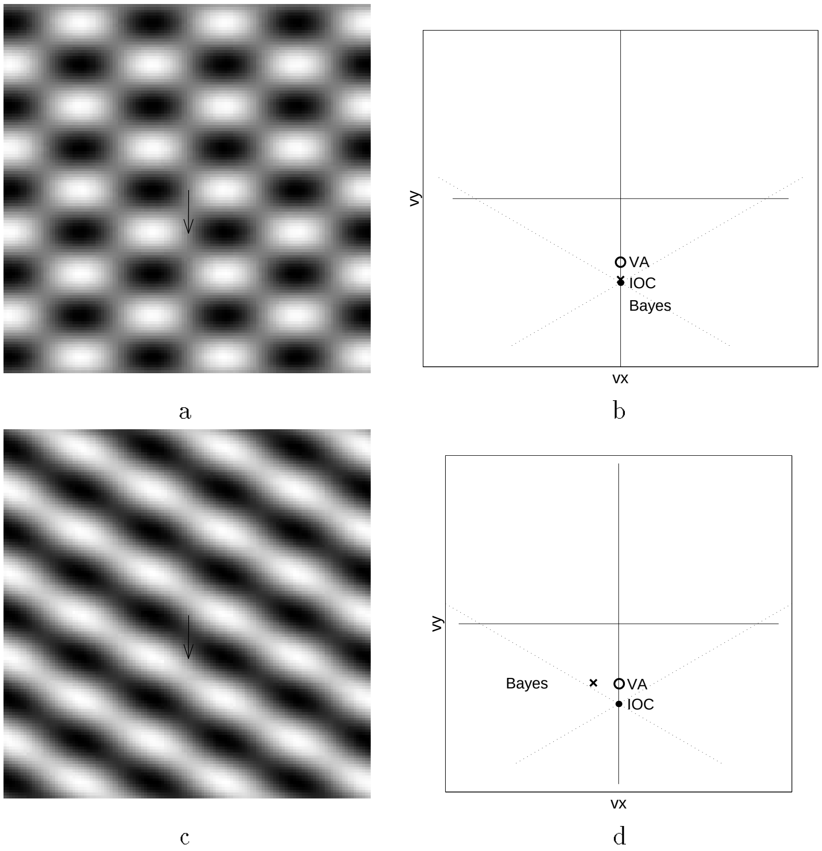


Figure 28: The influence of relative contrast on the perceived direction of a moving type I plaid [35]. When both components are of identical contrasts the perceived motion is in the veridical direction. When they are of unequal contrasts, the perceived direction is biased in the direction of the higher contrast grating. A similar pattern is observed in the output of the Bayesian estimator.

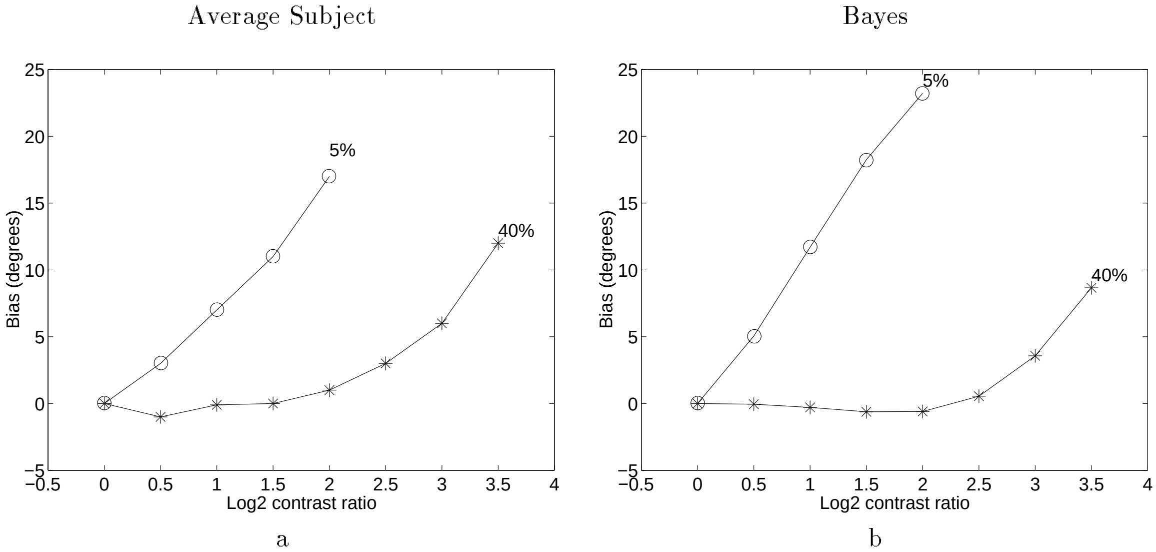


Figure 29: An experiment conducted by Stone et al. (1990) showing the influence of relative contrast on the perceived direction of a moving plaid. Subjects viewed a set of type I plaids and the contrasts of the two components was systematically varied. **a.** Results averaged over subjects replotted from. [35]. The direction of motion of the plaid was biased in the direction of the higher contrast grating and the magnitude of the bias decreases with increased total contrast. **b.** The Bayesian estimator gives similar results. (cf. [11]).

plaid, i.e. the sum of the contrasts of the two gratings. When the contrast of both gratings was increased (while the ratio of contrast stayed constant) a smaller bias was observed. Figure 29a shows data averaged over subjects replotted from [35].

Model Results: The results of the Bayesian estimator are shown in figure 29b. Similar to the results of human observers, the estimate is biased in the direction of the higher contrast grating and the magnitude of the bias decreases with increasing total contrast.

Discussion: Again this is a result of the fact that as contrast is decreased the local uncertainty decreases. Thus in figure 28d, the likelihood corresponding to the low contrast grating is a very "fuzzy" constraint line. In this case, although the Bayesian solution does not lie exactly on both constraint lines it has very similar likelihood to the IOC solution. In terms of the prior, however, the Bayesian solution is favored because it is slower. When both gratings are of identical contrasts, the likelihoods have equal fuzziness and the Bayesian solution has the correct direction (although the magnitude is smaller than the IOC solution). When the total contrast is increased, all the likelihoods become more peaked and the Bayesian solution is forced to lie closer to the IOC solution.

Although the results of the Bayesian estimator is in qualitative agreement with the psychophysical results for this task, the quantitative fit can be improved. Heeger and Simoncelli (1991) have obtained better fits for this data using their model that also includes a nonlinear gain control mechanism.

4.6 Dependence of bias on duration

4.6.1 Dependence of type II bias on duration - Yo and Wilson (1992)

Phenomena: Yo and Wilson (1992) reported that the perceived direction of type II plaids changes with stimulus duration. At short durations, the perceived direction is heavily biased in the direction of the vector average and gradually approaches the IOC solution as duration is increased. Figure 30b shows the results of a single subject.

Model Results: Figure 30c shows the predictions of the Bayesian estimator. The model was given 5 frames of the video sequence, and the local likelihood was calculated by summing filter outputs over space and time. In that respect the results in this section differ from those reported in other sections, where only two frames were used to calculate the local likelihoods. Note the change in model output with increased duration.

Discussion: As discussed in section 3.1, short durations serve to make the local likelihood less peaked. In fact, the short duration acts in the model much like low contrast (figure 25). At short durations, there is only a small difference between the degree to which the true velocity satisfies the gradient constraint and the degree to which other velocities do so. However, as gradient information is combined over time, the difference becomes more pronounced and the uncertainty in the local measurement decreases. The shorter the presentation time the more the local information is ambiguous.

While the sharpness of the local likelihood change with duration, the prior probability does not. Hence the VA solution which has a higher prior probability is favored at short durations, while at long durations the Bayesian estimate approaches the IOC solution.

4.6.2 Dependence on duration in line drawings – Lorenceau et al. 1992

Phenomena: Lorenceau et al. (1992) reported a similar effect of duration in the discrimination of line motion. As explained in the previous section, subjects were requested to judge whether the matrix of lines moved above or below the horizontal. At short durations, they found that performance was below chance, indicating that subjects perceived the lines moving in the normal direction, but performance improved at longer durations. Figure 31b shows the results of a single subject replotted from [15]. Despite significant individual variations, subjects consistently perform below chance at short durations and improve as duration increases.

Model Results: Figure 31c shows the output of the Bayesian estimator. A single mechanism predicts systematic errors at short durations with an increase in correct responses as duration is increased. Note that this explanation does not require separate "1D" and "terminator" mechanisms. Rather it is explained in the same way as the influence of duration on plaids.

Again, at low durations all local measurements have higher degree of uncertainty. In the Bayesian model there is no categorization of location into "1D" or "2D" but at all locations the gradient constraint is accumulated over space and time. At short durations, therefore, there is less signal in the local spatiotemporal window, and hence more uncertainty in the local likelihoods. In this condition, the prior favoring slow speeds dominates and perception is in the normal direction. At long durations, the local uncertainty is decreased, and the prior has a much weaker influence.

Discussion: The results reported in this section were obtained by using a spatiotemporal Gaussian window in equation 6. This gives an additional free parameter to fit the data. However the qualitative nature of the results are unchanged when the window function is changed. Any summation of information over time would lead to a decrease in local uncertainty with longer durations. Thus a Bayesian estimation strategy predicts highly biased estimates at low durations but more veridical velocity as duration increases.

4.7 Non-translational motions

So far we have discussed stimuli undergoing uniform translation. Although the model returns a flow field we

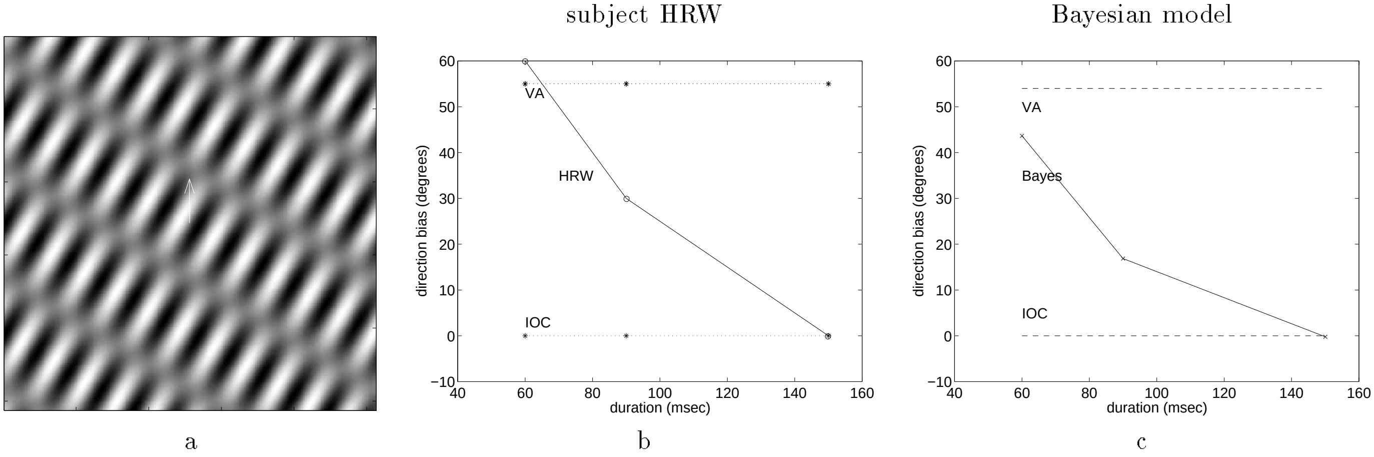


Figure 30: The influence of duration on performance in the experiment conducted by Yo and Wilson (1992). At short durations, the perceived motion is heavily biased towards the VA, and it approaches the IOC solutions at long durations. **a.** a single frame from the sequence. **b.** The results of subject HRW replotted from [45]. **c.** The predictions of a Bayesian estimator. The predicted velocity shows a gradual shift from VA to IOC as duration increases.

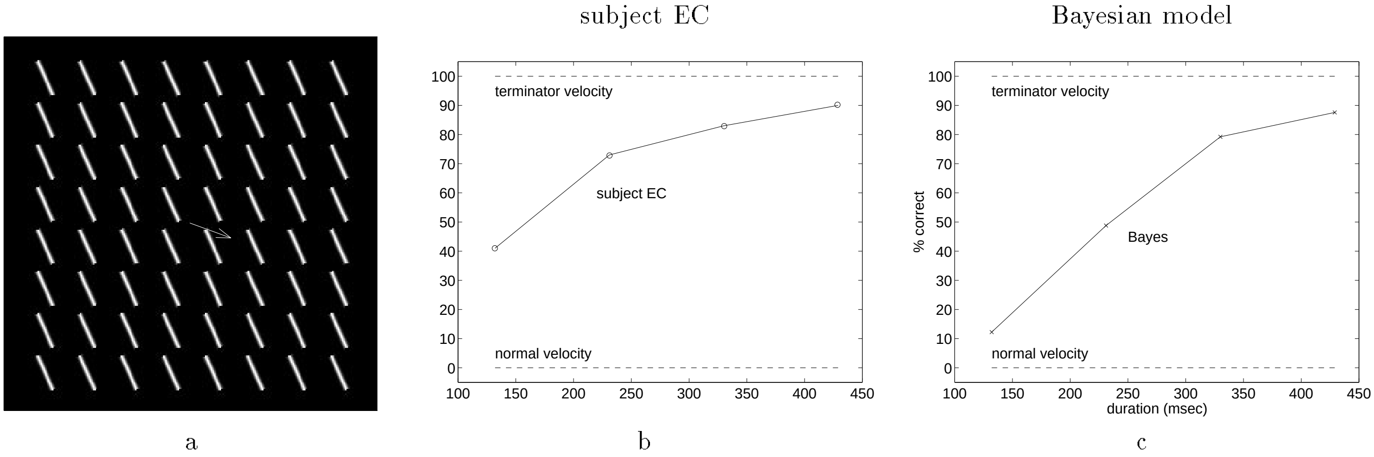


Figure 31: The influence on duration on performance in the experiment conducted by Lorenceau et al. (93). At short duration, performance is below chance indicating subjects perceive motion in the normal direction, while at long durations the perceived motion is largely veridical. **a.** a single frame from the sequence. **b.** The results of a single subject replotted from [15]. Despite significant individual variations, subjects consistently perform below chance at short durations and improve as duration increases. **c.** The predictions of a Bayesian estimator. A single Bayesian mechanism would predict systematic errors at short durations with an increase in correct responses as duration is increased.

could capture it with a single velocity vector. Now we show the output of the model on non-translational motions. We display the output of the model by plotting arrows at different (arbitrarily chosen) locations of the image.

4.7.1 Circles and derived figures in rotation - Wallach 1956

Phenomena: Musatti (1924) and Wallach et al. (1956) observed that when circular figures are rotated in the image plane (e.g by putting them on a turntable) they are not perceived as rigidly rotating. A rotating circle appears static, a rotating spiral appears to contract, and a rotating ellipse appears to deform nonrigidly. In the case of the rotating ellipse, Wallach et al. (1956) noted that the perceived rigidity is most pronounced when the ellipse is "fat" — with aspect ratio close to unity. Musatti pointed out that when a small number of rotating features are added to the display, the rigid percept becomes prominent.

Model Results: Figure 32 shows the output of the Bayesian estimator on these stimuli. As in human perception the rotating circle is perceived as static, the rotating spiral as expanding and the rotating ellipse as deforming nonrigidly. Figure 33 shows the model output on a narrow ellipse and on an ellipse with four rotating features added. Note that in this case, consistent with human perception, the predicted motion is much closer to rotation. The parameter σ is held constant.

Discussion: Why does the model "misperceive" these motions? First note that for the stimuli in figure 32, the perceived motions and the rotational motions have very similar likelihoods. That is, due to the low curvature of the figure, the local likelihoods are highly ambiguous. Given that the likelihoods are nearly identical, the Bayesian estimator is dominated by the prior. Here again, the "slowness" prior may be responsible for the percept. Figure 34 shows the total magnitude of the velocity fields. Note that the rotational velocity is much faster than the Bayesian estimate, and hence is not favored.

The Bayesian estimate considers both the likelihood and the prior. Thus once the rotating stimulus includes locations that are relatively ambiguous (e.g. the endpoints of a narrow ellipse, or dots flanking the fat ellipse), the estimate resembles rotation. The rotation still has lower prior probability but high likelihood.

A slightly different account of these illusions was given by Hildreth (1983). Her model chooses the velocity field of least variation that satisfies the gradient constraint at every location along the ellipse. Although her algorithm did not include an explicit penalty for fast velocity fields it gave similar results to those shown here – a rotating circle was estimated to be stationary, a rotating spiral was estimated to be expanding and a rotating fat ellipse was estimated to be deforming.

Note however that by penalizing the magnitude of the first derivative, Hildreth's algorithm includes an implicit penalty for fast non-translational velocity fields. That is, for all translational velocity fields, the first derivative is zero everywhere and there is no distinction between fast and slow fields. For velocity fields whose first derivative does not vanish, however, the magnitude of the first derivative increases with increased speed. Thus Hildreth's algorithm will in general prefer a slow deformation to a faster rotation. It will not, however, prefer a slow translation to a faster one, and thus can not account for biases encountered in translating stimuli (e.g. the VA bias in plaids).

4.7.2 Smooth curves in translation - Nakayama and Silverman 1988

Phenomena: Nakayama and Silverman (1988) found that smooth curves including sinusoids, Gaussians and sigmoids, may be perceived to deform nonrigidly when they are translated rigidly in the image plane. Figure 35a shows an example. A "shallow" sinusoid is translating rigidly horizontally. This stimulus is typically perceived as deforming nonrigidly. The authors noted that the perceived nonrigidity was most pronounced for "shallow" sinusoids in which the curvature of the curves was small.

Model Results: Figure 35b shows the output of the Bayesian estimator. For the shallow sinusoid the Bayesian estimator favors a slower hypothesis than the veridical rigid translation. Figure 35d shows the output on the sharp sinusoid. Note that a fixed σ gives a nonrigid percept for the shallow sinusoid and a rigid percept for the sharp sinusoid.

Discussion: Again this is the result of the tradeoff between "slow" and "smooth" priors. The nonrigid percept is slower than the rigid translation but less smooth. For shallow sinusoids, the nonrigid percept is still relatively smooth, but for sharp sinusoids the smoothness term causes the rigid percept to be preferred. The shape of sinusoid for which the percept will shift from rigid to nonrigid depends on the free parameter λ which governs the tradeoff between the slowness and smoothness terms. The qualitative results however remain the same — sharp sinusoids are perceived as more rigid than shallow ones. Similar results were also obtained with the other smooth curves studied by Nakayama and Silverman — the Gaussian and the sigmoidal curves.

5 Discussion

Since the visual system receives information that is ambiguous and uncertain, it must combine the input with prior constraints to achieve reliable estimates. A Bayesian estimator is the simplest reasonable approach and the prior favoring slow and smooth motions offer reasonable constraints. In this paper we have asked how such a system will behave. We find that, like humans, its motion estimates include apparent biases and illusions.

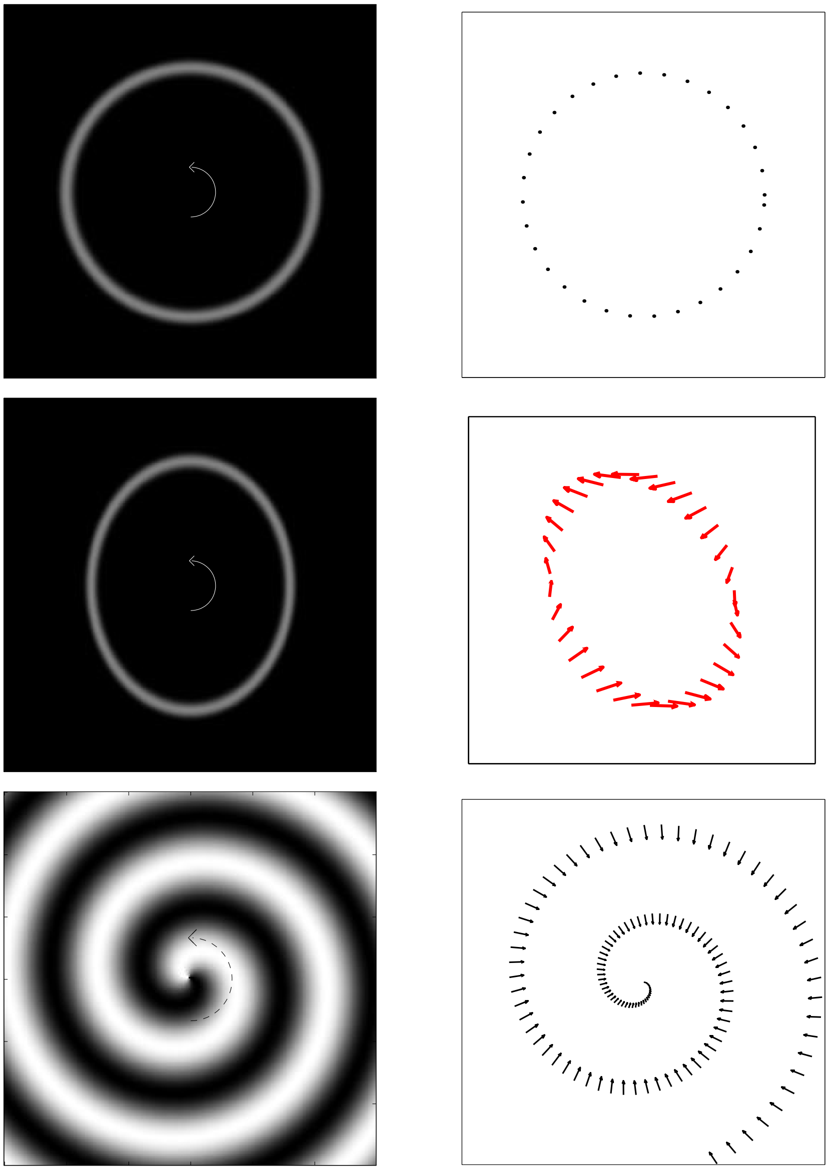


Figure 32: Biased perception in Bayesian estimation of circles and derived figures in rotations. Due to the prior favoring slow and smooth velocities, the estimate may be biased away from the veridical velocity and towards the normal components. These biases are illustrated here. A rotating circle appears to be stationary, a rotating ellipse appears to deform nonrigidly, and a rotating spiral appears to expand and contract.

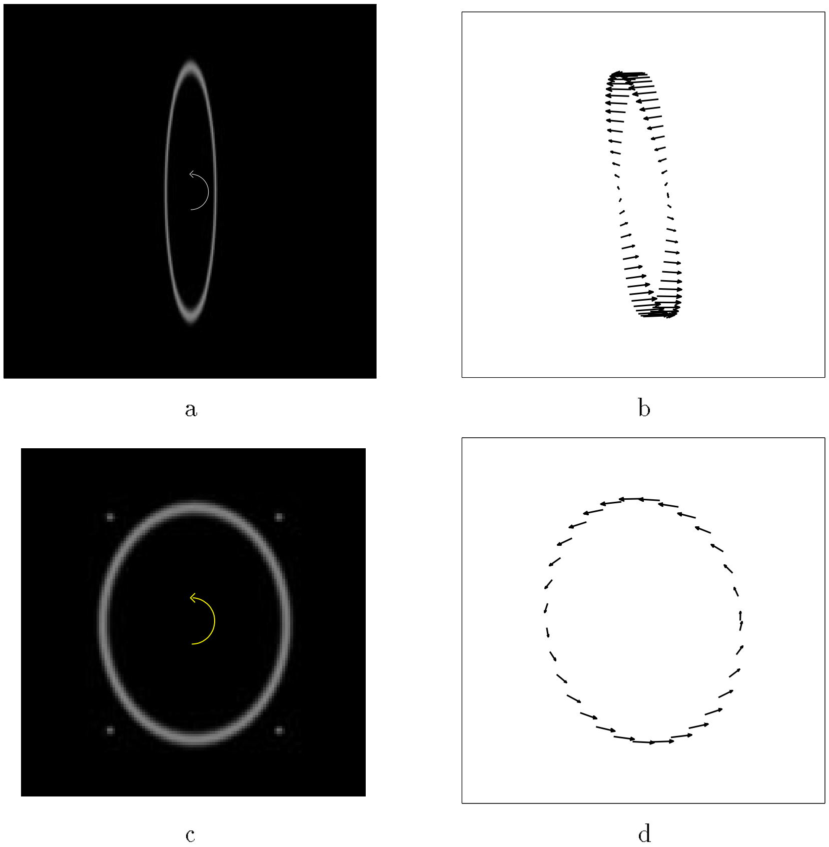


Figure 33: The percept of nonrigid deformation is influenced by stimulus shape and by additional features. For a “narrow” rotating ellipse, the Bayesian estimate is similar to rotation. Similarly, for a “fat” rotating ellipse with four rotating dots, the estimate is similar to rotation. This is consistent with human perception. The parameter σ is held constant.

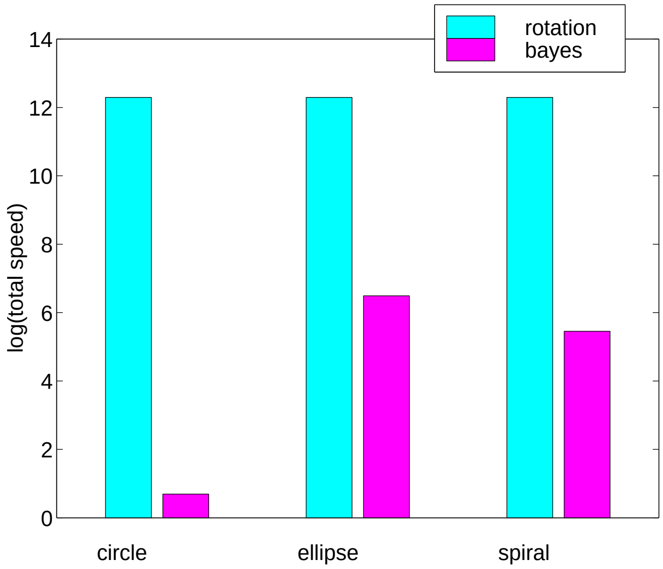


Figure 34: The total magnitude of the velocity fields arrived at by the Bayesian estimate for the stimuli in 32 as compared to the true rotation. Note that the rotational velocity is much faster than the Bayesian estimate, and hence is not favored.

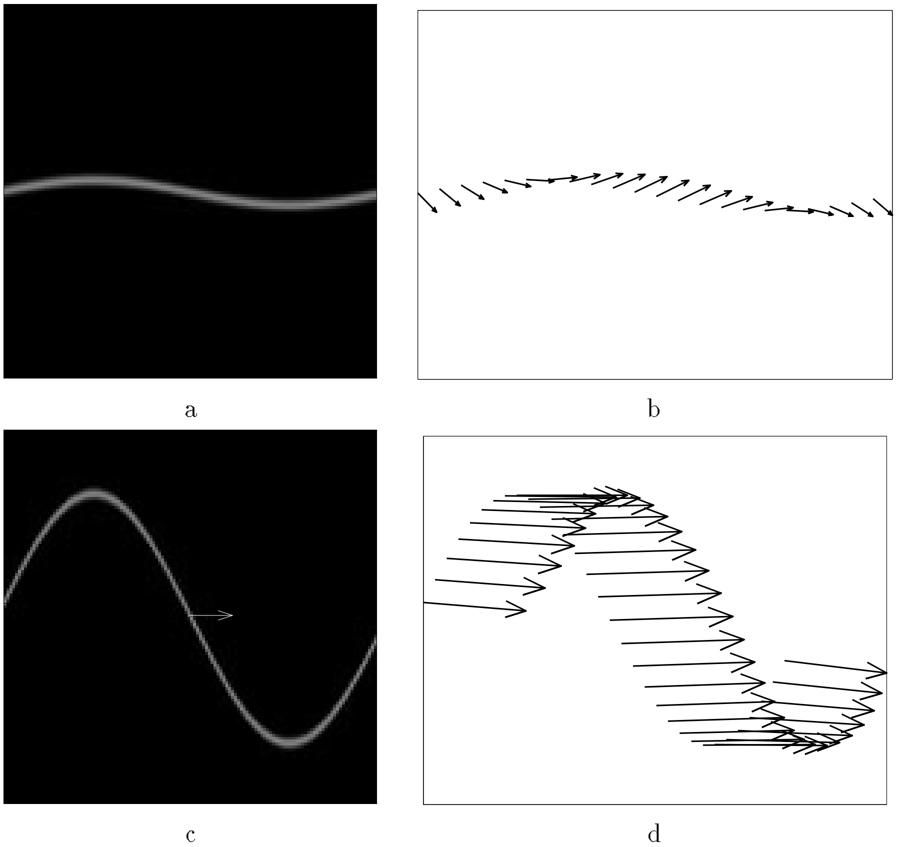


Figure 35: **a.** A “shallow” sinusoid translating horizontally appears to to deform nonrigidly (Nakayama and Silverman 1988). **b.** The nonrigid deformation is also prevalent in the Bayesian estimator. **c.** A “sharp” sinusoid translating horizontally appears to translate rigidly (Nakayama and Silverman 1988). **d.** Rigid translation is also prevalent in the Bayesian RBF estimator.

Moreover, this non-veridical perception is quite similar to that exhibited by humans in the same circumstances.

In recent years a large number of phenomena have been described in velocity estimation, usually connected with the aperture problem. In reviewing a long list of phenomena, we find that the Bayesian estimator almost always predicts the psychophysical results. The predictions agree qualitatively, and are often in remarkable agreement quantitatively.

The Bayesian estimator is a simple and reasonable starting point for a model of motion perception. Insofar as it explains the data, there is no need to propose specific mechanisms that deal with lines, terminators, plaids, blobs etc. These other mechanisms are often poorly defined, and they are often assumed to turn on or off according to special rules.

The Bayesian estimator described here can be applied to any image sequence that contains a single moving surface. It works with gratings, plaids, ellipses or spirals without modification. It usually needs only a single free parameter σ, which corresponds to the noise or internal uncertainty level in the observer's visual system. Even this parameter remains fixed when the individual observer and viewing conditions are fixed.

Beyond the specifics of our particular model, we have shown that human motion perception exhibits two fundamental properties of a Bayesian estimator. First, observers give different amounts of weight to information at different locations in the image - e.g. a small number of features can profoundly influence the percept and high contrast locations have greater influence than low contrast ones. This is consistent with a Bayesian mechanism that combines sources of evidence in accordance with their uncertainty. Second, the motion percept exhibits a bias towards slow and smooth velocities, consistent with a Bayesian mechanism that incorporates prior knowledge as well as evidence into the estimation.

Each of these properties have appeared in some form in previous models. The notion of giving unequal weight to different motion measurements appears, for example, in the model suggested by Lourenceau et al. (1992). Mingolla et al. (1992) suggested assigning these weights according to their "saliencies" which would in turn depend on contrast. In the Bayesian framework, the amount of weight given to a particular measurement has a concrete source — it depends on its uncertainty. Thus the low weight given to low contrast, short duration or peripherally viewed features is a consequence of the high degree of uncertainty associated with them. Moreover, there is no need to arbitrarily distinguish between "2D" and "1D" local features — all image regions have varying degrees of uncertainty, and the strong influence of cornerlike features is a consequence of the relatively unambiguous motion signals they give rise to.

As mentioned in the introduction, the models of Hildreth (1983) and Grzywacz and Yuille (1991) include a bias towards smooth velocity fields. However these algorithms do not have the concept of varying degrees of ambiguity in local motion measurements. They either represents the local information as a constraint line in velocity space or as a completely unambiguous 2D measurement. They therefore can not account for the gradual shift in perceived direction of figures as contrast and duration are varied.

The smoothness assumption used by Hildreth (1983) and others, can be considered a special case of the regularization approach to computational vision introduced by Poggio et al. (1985). This approach is built on the observation that many problems in vision are "ill-posed" in the mathematical sense — there are not enough constraints in the data to reliably estimate the solution. Regularization theory [37] provides a general mathematical framework for solving such ill-posed problems by minimzing cost functions that are the sum of two terms – a "data" term and a "regularizer" term. There are very close links between Bayesian MAP estimation and regularization theory (e.g. [18]). In the appendix we show how the Bayesian motion theory presented here could be rephrased in terms of regularization theory.

The model of Heeger and Simoncelli (1991) was to the best of our knowledge, the first to provide a Bayesian account of human motion perception that incorporatea a prior favoring slow speeds. Indeed the first stage of our model, the extraction of local likelihoods, is very similar to the Heeger and Simoncelli model. In our model, however, these local likelihoods are then combined across space to estimate a spatially varying velocity field. In spatially isotropic stimuli (such as plaids and gratings) there is no need to combine across space as all spatial locations give the same information. However, integration across space is crucial in order to account for motion perception in more general stimuli such as translating rhombuses, rotating spirals or translating sinusoids.

Another local motion analysis model was introuced by Bulthoff et al. (1989) who described a simple, parallel algorithm that computes optical flow by summing activities over a small neighborhood of the image. Unlike the Heeger and Simoncelli model, their model did not include a prior favoring slow velocities and therefore predicts the IOC solution for all plaid stimuli.

We have attempted to make the Bayesian estimator discussed here as simple as possible, at the sacrifice of biological faithfulness. Thus we assume a Gaussian noise model, a fixed σ and linear gradient filters. One disadvantage of this simple model is that in order to obtain quantitative fits to the results of existing experiments we had to vary σ between experiments (but σ was always held fixed when modeling a single experiment with multiple conditions). Although changing σ does not in general change the qualitative nature of the Bayesian estimate, it does change the quantitative results. A more complicated Bayesian estimator, that also models the nonlinearities in early vision, may be able to fit more data with fixed parameters.

How could a Bayesian estimator of the type discussed here be implemented given what is known about the functional architecture of the primate visual system? The local likelihoods are simple functions (squaring and summing) of the outputs of spatiotemporal filters at a particular location. Thus a population of units in primary visual cortex may be capable of representing these local likelihoods [11]. Combining the likelihoods and finding the most probable velocity estimate, however, is a more complicated matter and is an intriguing question for future research.

Indeed understanding the mechanism by which human vision combines local motion signals may prove fruitful in the design of artificial vision systems. Human motion perception seems to accurately represent uncertainty of local measurements, and to combine these measurements in accordance with their uncertainty together with a prior probability. Despite this sophistication motion perception is immediate and effortless, suggesting that the human visual system has found a way to perform fast Bayesian inference.

Acknowledgments

We thank D. Heeger, E. Simoncelli, N. Rubin, J. McDermott, and H. Farid for discussions and comments.

References

[1] Edward H. Adelson and James R. Bergen. The extraction of spatio-temporal energy in human and machine vision. In *Proceedings of the Workshop on Motion: Representation and Analysis*, pages 151–155, Charleston, SC, 1986.

[2] E.H. Adelson and J.A. Movshon. Phenomenal coherence of moving visual patterns. *Nature*, 300:523–525, 1982.

[3] D. Alais, P. Wenderoth, and D. Burke. The contribution of one-dimensional motion mechanisms to the perceived direction of drifting plaids and their aftereffects. *Vision Research*, 34(14):1823–1834, 1994.

[4] L. Bowns. Evidence for a feature tracking explanation of why type II plaids move in the vector sum directions at short directions. *Vision Research*, 36(22):3685–3694, 1996.

[5] O. Braddick. Segmentation versus integration in visual motion processing. *Trends in Neuroscience*, 16:263–268, 1993.

[6] H. Bulthoff, J. Little, and T. Poggio. A parallel algorithm for real-time computation of optical flow. *Nature*, 337(6207):549–553, 1989.

[7] Darren Burke and Peter Wenderoth. The effect of interactions between one-dimensional component gratings on two dimensional motion perception. *Vision Research*, 33(3):343–350, 1993.

[8] V.P. Ferrera and H.R. Wilson. Perceived direction of moving two-dimensional patterns. *Vision Research*, 30:273–287, 1990.

[9] Federico Girosi, Michael Jones, and Tomaso Poggio. Regularization theory and neural networks architectures. *Neural Computation*, 7:219–269, 1995.

[10] N.M. Grzywacz and A.L. Yuille. Theories for the visual perception of local velocity and coherent motion. In J. Landy and J. Movshon, editors, *Computational models of visual processing.* MIT Press, Cambridge, Massachusetts, 1991.

[11] David J. Heeger and Eero P. Simoncelli. Model of visual motion sensing. In L. Harris and M. Jenkin, editors, *Spatial Vision in Humans and Robots.* Cambridge University Press, 1991.

[12] E. C. Hildreth. *The Measurement of Visual Motion.* MIT Press, 1983.

[13] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artif. Intell.*, 17(1–3):185–203, August 1981.

[14] D. Knill and W. Richards. *Perception as Bayesian Inference.* Cambridge University Press, 1996.

[15] Jean Lourenceau, Maggie Shifrar, Nora Wells, and Eric Castet. Different motion sensitive units are involved in recovering the direction of moving lines. *Vision Research*, 33(9):1207–1217, 1992.

[16] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Image Understanding Workshop*, pages 121–130, 1981.

[17] D. Marr and S. Ullman. Directional selectivity and its use in early visual processing. *Proceedings of the Royal Society of London B*, 211:151–180, 1981.

[18] J.L. Marroquin, S. Mitter, and T. Poggio. Probabilistic solution of ill-posed problems in computational vision. *Journal of the American Statistical Association*, 82:76–89, 1987.

[19] E. Mingolla, J.T. Todd, and J.F. Norman. The perception of globally coherent motion. *Vision Research*, 32(6):1015–1031, 1992.

[20] A.J. Movshon, E.H. Adelson, M.S. Gizzi, and W.T. Newsome. The analysis of moving visual patterns. *Experimental Brain Research*, 11:117–152, 1986.

[21] C.L. Musatti. Sui fenomeni stereocinetici. *Archivio Italiano di Psicologia*, 3:105–120, 1924.

[22] C.L. Musatti. Stereokinetic phenomena and their interpretation. In Giovanni B. Flores Darcais, editor, *Studies in Perception: Festschrift for Fabio Metelli.* Martello - Giunti, Milano, 1975.

[23] K. Nakayama and G. H. Silverman. The aperture problem - I: Perception of nonrigidity and motion direction in translating sinusoidal lines. *Vision Research*, 28:739–746, 1988.

[24] K. Nakayama and G. H. Silverman. The aperture problem - II: Spatial integration of velocity information along contours. *Vision Research*, 28:747–753, 1988.

[25] Steven J. Nowlan and Terrence J. Sejnowski. A selection model for motion processing in area MT of primates. *The Journal of Neuroscience*, 15(2):1195–1214, 1995.

[26] T. Poggio and W. Reichardt. Considerations on models of movement detection. *Kybernetik*, (13):223–227, 1973.

[27] T. Poggio, V. Torre, and C. Koch. Computational vision and regularization theory. *Nature*, 317:314–319, 1985.

[28] W. Reichardt. Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In W. A. Rosenblith, editor, *Sensory Communication.* Wiley, 1961.

[29] H.R. Rodman and T.D. Albright. Single-unit analysis of pattern motion selective properties in the middle temporal visual area MT. *Experimental Brain Research*, 75:53–64, 1989.

[30] Nava Rubin and Saul Hochstein. Isolating the effect of one-dimensional motion signals on the perceived direction of moving two-dimensional objects. *Vision Research*, 33:1385–1396, 1993.

[31] S. Shimojo, G. Silverman, and K. Nakayama. Occlusion and the solution to the aperture problem for motion. *Vision Research*, 29:619–626, 1989.

[32] E. P. Simoncelli. *Distributed Representation and Analysis of Visual Motion.* PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts of Technology, Cambridge, January 1993.

[33] E.P. Simoncelli and D.J. Heeger. A computational model for perception of two-dimensional pattern velocities. *Investigative Opthamology and Vision Research*, 33, 1992.

[34] E.P. Simoncelli and D.J. Heeger. A model of neuronal responses in visual area MT. *Vision Research*, 38(5):743–761, 1998.

[35] L.S. Stone, A.B. Watson, and J.B. Mulligan. Effect of contrast on the perceived direction of a moving plaid. *Vision Research*, 30(7):1049–1067, 1990.

[36] P. Thompson, L.S. Stone, and S. Swash. Speed estimates from grating patches are not contrast normalized. *Vision Research*, 36(5):667–674, 1996.

[37] A.N. Tikhonov and V.Y. Arsenin. *Solution of Ill-Posed problems.* W.H. Winston, Washington DC, 1977.

[38] Shimon Ullman. *The interpretation of visual motion.* The MIT Press, 1979.

[39] H. Wallach. Ueber visuell whargenommene bewegungrichtung. *Psychologische Forschung*, 20:325–380, 1935.

[40] H. Wallach, A. Weisz, and P. A. Adams. Circles and derived figures in rotation. *American Journal of Psychology*, 69:48–59, 1956.

[41] Y. Weiss. Smoothness in layers: Motion segmentation using nonparametric mixture estimation. In *Proceedings of IEEE conference on Computer Vision and Pattern Recognition*, pages 520–527, 1997.

[42] L. Welch. The perception of moving plaids reveals two processing stages. *Nature*, 337:734–736, 1989.

[43] H.R. Wilson, V.P. Ferrera, and C. Yo. A psychophysically motivated model for two-dimensional motion perception. *Visual Neuroscience*, 9:79–97, 1992.

[44] S. Wuerger, R. Shapley, and N. Rubin. On the visually perceived direction of motion by hans wallach: 60 years later. *Perception*, 25:1317–1367, 1996.

[45] C. Yo and H.R. Wilson. Perceived direction of moving two-dimensional patterns depends on duration, contrast, and eccentricity. *Vision Research*, 32(1):135–147, 1992.

Appendix

5.1 Solving for the most probable velocity field

We derive here the equations for finding the parametric vector that maximizes the posterior probability. To simplify the notation, we denote the location (x, y) with a single vector r. Assume that the velocity field $v(r)$ is composed of a sum of N basis functions with the coefficients defined by the parameter vector θ. Define $\Psi(r)$ a 2 by N matrix which give the two components of the basis functions at location r, then $v(r) = \Psi(r)\theta$. Using this notation we can now rewrite the likelihoods and the prior as a function of θ.

Recall that the local likelihood is given by:

$$L_r(v) = \alpha e^{-\sum_r w(r)(I_x v_x + I_y v_y + I_t)^2/2\sigma^2} \tag{14}$$

(we use the convention that for any probability distribution α represents the normalization constant that guarantees that the distribution sum to unity). By completing the square, this can be rewritten:

$$L_r(v) = \alpha e^{-(v-\mu(r))^t \Sigma^{-1}(r)(v-\mu(r))/2\sigma^2} \tag{15}$$

where $\mu(r), \Sigma^{-1}(r)$ represent the mean and covariance matrices of the local likelihood.

$$\Sigma^{-1}(r) = \sum_s w_{rs} \begin{pmatrix} I_x^2(s) & I_x(s)I_y(s) \\ I_x(s)I_y(s) & I_y^2(s) \end{pmatrix} \tag{16}$$

and $\mu(r)$ a solution to:

$$\Sigma^{-1}(r)\mu(r) = y(r) \tag{17}$$

with

$$y(r) = \sum_s w_{rs} \begin{pmatrix} I_x(s)I_t(s) \\ I_y(s)I_t(s) \end{pmatrix} \tag{18}$$

Substituting $v(r) = \Psi(r)\theta$ into equation 15 gives the local likelihood of the image derivatives given θ:

$$L_r(\theta) = \alpha e^{-(\Psi(r)\theta-\mu(r))^t \Sigma^{-1}(r)(\Psi(r)\theta-\mu(r))/2\sigma^2} \tag{19}$$

and finally assuming conditional independence, the global likelihood for the image derivatives is given the product of the local likelihoods at all locations:

$$L(\theta) = \Pi_r L_r(\theta) \tag{20}$$

We now express the prior probability as a function of θ. Recall that the prior favors slow and smooth velocities:

$$P(V) = \alpha e^{-\sum_r (Dv)^t(r)(Dv)(r))/2} \tag{21}$$

where D is a differential operator. Substituting $v(r) = \Psi(r)\theta$ gives the prior probability on θ:

$$P(\theta) = \alpha e^{-\theta^t R\theta/2} \tag{22}$$

Where R is a symmetric, NxN matrix such that

$$R_{ij} = \sum_r (D\Psi_i^t)(r)(D\Psi_j)(r) \tag{23}$$

where we have used $\Psi_i(r)$ the ith basis field, and $D\Psi_i(r)$ the results of applying the differential operator D to that basis field.

The posterior is given by:

$$P(\theta|I) = \alpha P(\theta)P(I|\theta) \tag{24}$$

The log-posterior is given by:

$$\log P(\theta|I) = k - \theta^t R\theta/2\sigma_p^2 \tag{25}$$
$$+\sum_r -(\Psi(r)\theta - \mu(r))^t \Sigma^{-1}(r)(\Psi(r)\theta - \mu(r))/2\sigma^2$$

(note that the log-posterior is quadratic in θ or in other words the posterior is a Gaussian distribution. Thus maximizing the posterior is equivalent to taking its mean)

To find θ^* the value of θ that maximizes the posterior we solve:

$$A\theta^* = b \tag{26}$$

with:

$$A = \left(\sum_r \Psi^t(r)\Sigma^{-1}(r)\Psi(r)/\sigma^2 + R/\sigma_p^2\right) \tag{27}$$

$$b == \left(\sum_r \Psi^t(r)\Sigma^{-1}\mu(r)\right)/\sigma^2 \tag{28}$$

Specifically, the parameters we use in these simulations are as follows. The differential operator D was chosen so that the Green's functions corresponding to it were Gaussians with standard deviation equal to 70% of the size of the image. The basis fields were also Gaussians with the same standard deviations. We used 50 basis fields, 25 with purely horizontal velocity and 25 with pure vertical velocity. The centers of the basis fields were equally spaced in the image, i.e. were placed on a $5x5$ grid. In this case the matrix R has a particularly simple form. If Ψ_i and Ψ_j are both vertical (or horizontal) then R_{ij} is simply the value of the ith basis field evaluated at the center of the jth basis field. Otherwise, $R_{ij} = 0$.

To summarize, given an image sequence and a parameterization of the velocity field, the Bayesian estimate of motion is obtained by solving equation 26. Finally the optimal velocity field is obtained by $v(r) = \Psi(r)\theta^*$.

5.2 Relation to regularization theory

There are very close links between Bayesian MAP estimation and regularization theory (e.g. [18]). For completeness, we now show how to rephrase the Bayesian motion theory presented here in terms of regularization theory.

Regularization theory calls for minimizing cost functions that have two terms: a "data" term and a "regularizer term". A classical example is function approximation where one is given samples $\{x_i, y_i\}$ and wishes to find the approximating function. Obviously this is an ill-posed problem – there are an infinite number of functions that could approximate the data equally well. A typical regularization approach calls for minimizing:

$$J(f) = \sum_i (f(x_i) - y_i)^2 + \lambda \int_x \|Df(x)\|^2 dx \qquad (29)$$

The first term on the right hand side is the data term and the second term is the regularizer, in this case regularization is performed by penalizing for high derivatives.

Note that the log posterior in equation 25 can also be decomposed into two terms that depend on θ . The sum of the log likelihoods $\sum_r -(\Psi(r)\theta - \mu(r))^t \Sigma^{-1}(r)(\Psi(r)\theta - \mu(r))/2\sigma^2$ and the log prior $-\theta^t R\theta/2\sigma_p^2$. In the language of regularization theory, the negative sum of the log likelihoods would be the "data term" and the negative log posterior would be the "regularizer term".

The negative log posterior, when considered as a "regularizer" is quite similar to the smoothness regularizer in equation 29 in that it penalizes for values of θ that correspond to velocity fields that have large derivatives. Likewise the negative log likelihood is similar to the data term in equation 29 in that it penalizes for the squared error between the observed data and the predicted velocity field. The main difference, however, is that different observations are given different weights in the log posterior. Recall from section 2 that in Bayesian MAP estimation for Gaussian likelihoods the weight of an observation is inversely proportional to its variance, hence the Σ^{-1} factor in equation 25. Although the regularization framework is broad enough to encompass nonuniform weights for the data, it does not give a prescription for how to choose the weights.

An elegant result that can be derived in the regularization framework shows that the function f that minimizes J in equation 29 can be expressed as a superposition of basis functions (see [9] and references within). In contrast, here we assume a particular representation for the velocity field rather than deriving it. We do this because the number of basis functions required for the optimal function f is equal to the number of datapoints. In the case of motion analysis, this number is prohibitively large. For computational efficiency we prefer a low dimensional representation. We have found that as long as one uses the prior over velocity fields, the exact form of the representation used is not crucial — very similar results are obtained with different represenations [41].