Article

Probabilistic Motion Estimation Based on Temporal Coherence

BURGI, Pierre-Yves, YUILLE, Alan L., GRZYWACZ, Norberto M.

Abstract

We develop a theory for the temporal integration of visual motion motivated by psychophysical experiments. The theory proposes that input data are temporally grouped and used to predict and estimate the motion flows in the image sequence. This temporal grouping can be considered a generalization of the data association techniques that engineers use to study motion sequences. Our temporal grouping theory is epxressed in terms of the Bayesian generalization of standard Kalman flitering. To implement the theory, we derive a parallel network that shares some properties of cortical networks. Computer simulations of this network demonstrate that our theory qualitatively accounts for psychophysical experiments on motion occlusion and motion outliers. In deriving our theory, we assumed spatial factorizability of the probability distributions and made the approximation of updating the marginal distributions of velocity at each point. This allowed us to perform local computations and simplified our implementation. We argue that these approximations are suitable for the stimuli we are considering (for which spatial coherence effects [...]

Reference

BURGI, Pierre-Yves, YUILLE, Alan L., GRZYWACZ, Norberto M. Probabilistic Motion Estimation Based on Temporal Coherence. *Neural Computation*, 2000, vol. 12, no. 8, p. 1839-1867

DOI : 10.1162/089976600300015169

LETTER Communicated by Steven Nowlan

Probabilistic Motion Estimation Based on Temporal Coherence

Pierre-Yves Burgi
Centre Suisse d'Electronique et Microtechnique, 2007 Neuchâtel, Switzerland

Alan L. Yuille
Norberto M. Grzywacz
Smith-Kettlewell Eye Research Institute, San Francisco, CA 94115, U.S.A.

We develop a theory for the temporal integration of visual motion motivated by psychophysical experiments. The theory proposes that input data are temporally grouped and used to predict and estimate the motion flows in the image sequence. This temporal grouping can be considered a generalization of the data association techniques that engineers use to study motion sequences. Our temporal grouping theory is expressed in terms of the Bayesian generalization of standard Kalman filtering. To implement the theory, we derive a parallel network that shares some properties of cortical networks. Computer simulations of this network demonstrate that our theory qualitatively accounts for psychophysical experiments on motion occlusion and motion outliers. In deriving our theory, we assumed spatial factorizability of the probability distributions and made the approximation of updating the marginal distributions of velocity at each point. This allowed us to perform local computations and simplified our implementation. We argue that these approximations are suitable for the stimuli we are considering (for which spatial coherence effects are negligible).

1 Introduction

Local motion signals are often ambiguous, and many important motion phenomena can be explained by hypothesizing that the human visual system uses temporal coherence to resolve ambiguous inputs. (Temporal coherence is the assumption that motion in natural images is mostly temporally smooth, that is, it rarely changes direction or speed abruptly.) For example, the perceptual tendency to disambiguate the trajectory of an ambiguous motion path by using the time average of its past motion was first reported by Anstis and Ramachandran (Ramachandran & Anstis, 1983; Anstis & Ramachandran, 1987). They called this phenomenon *motion inertia*. Other motion phenomena involving temporal coherence include the improvement of velocity estimation over time (McKee, Silverman, & Nakayama, 1986; Ascher, 1998), blur removal (Burr, Ross, & Morrone, 1986), motion-outlier

detection (Watamaniuk, McKee, & Grzywacz, 1994), and motion occlusion (Watamaniuk & McKee, 1995). Another example of the integration of motion signals over time comes from a recent experiment by Nishida and Johnston (1999) on motion aftereffects, where an orientation shift in the direction of the illusory rotation appears to involve dynamic updating by motion neurons. These motion phenomena pose a serious challenge to motion perception models. For example, in motion-outlier detection, the target dot is indistinguishable from the background noise if one observes only a few frames of the motion. Therefore, the only way to detect the target dot is by means of its extended motion over time. Consequently, explanations based solely on single, large local-motion detectors as in the motion energy and elaborated Reichardt models (Hassenstein & Reichardt, 1956; van Santen & Sperling, 1984; Adelson & Bergen, 1985; Watson & Ahumada, 1985; for a review of motion models, see Nakayama, 1985) seem inadequate to explain these phenomena (Verghese, Watamaniuk, McKee, & Grzywacz, 1999). Also, it has been argued (Yuille & Grzywacz, 1998) that all these experiments could be interpreted in terms of temporal grouping of motion signals involving prediction, observation, and estimation. The detection of targets and their temporal grouping could be achieved by verifying that the observations were consistent with the motion predictions. Conversely, failure in these predictions would indicate that the observations were due to noise, or distractors, which could be ignored. The ability of a system to predict the target's motions thus represents a powerful mechanism to extract targets from background distractors and might be a powerful cue for the interpretation of visual scenes.

There has been little theoretical work on temporal coherence. Grzywacz, Smith, and Yuille (1989) developed a theoretical model that could account for motion inertia phenomena by requiring that the direction of motion varies slowly over time while speed could vary considerably faster. More recently Grzywacz, Watamaniuk, and McKee (1995) proposed a biologically plausible model that extends the work on motion inertia and allows for the detection of motion outliers. It was observed (Grzywacz et al., 1989) that for general three-dimensional motion, the direction, but not the speed, of the image motion is more likely to be constant and hence might be more reliable for motion prediction. This is consistent with the psychophysical experiments on motion inertia (Anstis & Ramachandran, 1987) and temporal smoothing of motion signals (Gottsdanker, 1956; Snowden & Braddick, 1989; Werkhoven, Snippe, & Toet, 1992), which demonstrated that the direction of motion is the primary cue in temporal motion coherence. Other reasons that might make motion direction more reliable than speed are the unreliability of local speed measurement due to the poor temporal resolution of local motion units (Kulikowski & Tolhurst, 1973; Holub & Morton-Gibson, 1981) and involuntary eye movements.

This article presents a new theory for visual motion estimation and prediction that exploits temporal information. This theory builds on our pre-

vious work (Grzywacz et al., 1989, 1995; Yuille & Grzywacz, 1998) and on recent work on tracking of object boundaries (Isard & Blake, 1996). In addition, this article gives a concrete example of the abstract arguments on temporal grouping proposed by Yuille and Grzywacz (1998). To this end, we use a Bayesian formulation of estimation over time (Ho & Lee, 1964), which allows us simultaneously to make predictions over time, update those predictions using new data, and reject data that are inconsistent with the predictions. (We will be updating the marginal distributions of velocity at each image point rather than the full probability distribution of the entire velocity field, which would require spatial coupling.) This rejection of data allows the theory to implement basic temporal grouping (for speculations about more complex temporal grouping, see Yuille & Grzywacz, 1998). This basic form of temporal grouping is related to data association as studied by engineers (Bar-Shalom & Fortmann, 1988) but differs by being probabilistic (following Ho & Lee, 1964). Finally, we show that our temporal grouping theory can be implemented using locally connected networks, which are suggestive of cortical structures.

In deriving our theory, we made an assumption of spatial factorizability of the probability distributions and made the approximation of updating the marginal distributions of velocity at each point. This allowed us to perform local computations and simplified our implementation. We argued that these assumptions and approximations (as embodied in our choice of prior probability) are suitable for the stimuli we are considering but would need to be revised to include spatial coherence effects (Yuille & Grzywacz, 1988, 1989). The prior distribution used in this article would need to be modified to incorporate these effects. (For speculations about how this might be achieved, see Yuille & Grzywacz, 1998.) There is no difficulty in generalizing the Bayesian theory to deal with complex hierarchical motion, but it is unclear whether the specific implementation in this article can be generalized.

The article is organized as follows. The theoretical framework for motion prediction and estimation is introduced in section 2. Section 3 gives a description of the specific model chosen to implement the theory in the continuous time domain. Implementation of the model using a locally connected network is addressed in sections 4 and 5. The model's predictions are tested in section 6 by performing computer simulations on the motion phenomena described previously. Finally, we discuss relevant issues, such as the possible relationship of the model to biology, in section 7.

2 The Theoretical Framework

2.1 Background. The theory of stochastic processes gives us a mathematical framework for modeling motion signals that vary over time. These processes can be used to predict the probabilities of future states of the system (e.g., the future velocities) and are called Markov if the probability of future states depends on only their present state (i.e., *not* on the time history

of how they arrived at this state). In computer vision, Markov processes have been used to model temporal coherence for applications such as the optimal fusing of data from multiple frames of measurements (Matthies, Kanade, & Szeliski, 1989; Clark & Yuille, 1990; Chin, Karl, & Willsky, 1994). Such optimal fusing is typically defined in terms of least-squares estimates, which reduces to Kalman filtering theory (Kalman, 1960). Kalman filters have been applied to a range of problems in computer vision (see Blake & Yuille, 1992, for several examples), and to neuronal mode by Rao and Ballard (1997). Because Kalman filters are (recursive) linear estimators that apply only to gaussian densities, their applicability in complex scenes involving several moving objects is questionable. One solution, which we briefly discuss, is the use of data association techniques (Bar-Shalom & Fortmann, 1988). In this article, however, we follow Isard and Blake (1996) and propose a Bayesian formulation of temporal coherence, which is a generalization of standard Kalman filters and can deal with targets moving in complex backgrounds.

We begin by deriving a generalization of Kalman filters that is suitable for describing the temporal coherence of simple motion. (Our development follows the work of Ho & Lee, 1964.) Consider the state vector $\vec{x}_k$ describing the status of a system at time t_k (in our theory, $\vec{x}_k$ denotes the velocity field at every point in the image). Our knowledge about how the system might evolve to time $k+1$ is described in a probability distribution function $p(\vec{x}_{k+1}|\vec{x}_k)$, known as the "prior" (in our theory, this prior will encode the temporal coherence assumption). The measurement process that relates the observation $\vec{z}_k$ of the state vector to its "true" state is described by the likelihood distribution function $p(\vec{z}_k|\vec{x}_k)$ (in our theory, $\vec{z}_k$ will be the responses of basic motion units at every place in the image). From these two distribution functions, it is straightforward to derive the a posteriori distribution function $p(\vec{x}_{k+1}|Z_{k+1})$, which is the distribution of $\vec{x}_{k+1}$ given the whole past set of measurements $Z_{k+1} \doteq (\vec{z}_0, \ldots, \vec{z}_{k+1})$ (note that although the prediction of the future depends only on the current state, the estimate of the current state does depend on the entire past history of the measurements). Using Bayes' rule and a few algebraic manipulations (Ho & Lee, 1964), we get

$$p(\vec{x}_{k+1}|Z_{k+1}) = \frac{p(\vec{z}_{k+1}|\vec{x}_{k+1})}{p(\vec{z}_{k+1}|Z_k)} p(\vec{x}_{k+1}|Z_k), \tag{2.1}$$

where $p(\vec{x}_{k+1}|Z_k)$ is prediction for the future state $\vec{x}_{k+1}$ given the current set of past measurements Z_k, and $p(\vec{z}_{k+1}|Z_k)$ is a normalizing term denoting the confidence in the measure $\vec{z}_{k+1}$ given the set of past measurements Z_k. The predictive distribution function can be expressed as

$$p(\vec{x}_{k+1}|Z_k) = \int p(\vec{x}_{k+1}|\vec{x}_k) p(\vec{x}_k|Z_k) d\vec{x}_k. \tag{2.2}$$

Equations 2.1 and 2.2 are generalizations of the Wiener-Kalman solution for linear estimation in presence of noise. Its evaluation involves two

stages. In the *prediction stage*, given by equation 2.2, the probability distribution $p(\vec{x}_{k+1}|Z_k)$ for the state at time $k+1$ is determined. This function, which involves the present state and the Markov transition describing the dynamics of the system, has a variance larger than that of $p(\vec{x}_k|Z_k)$. This increase in the variance results from the nondeterministic nature of the motion. In the *measurement stage*, performed by equation 2.1, the new measurements $\vec{z}_{k+1}$ are combined using Bayes' theorem and, if consistent, reinforce the prediction and decrease the uncertainty about the new state. (Inconsistent measurements may increase the uncertainty still further.)

It was shown (Ho & Lee, 1964) that if the measurement and prior probabilities are both gaussian, then equations 2.1 and 2.2 reduce to the standard Kalman equations, which update the mean and the covariance of $p(\vec{x}_{k+1}|Z_{k+1})$ and $p(\vec{x}_{k+1}|Z_k)$ over time. Gaussian distributions, however, are nonrobust (Huber, 1981), and an incorrect (outlier) measurement can seriously distort the estimate of the true state. Standard linear Kalman filter models are therefore *not able* to account for the psychophysical data that demonstrate, for example, that human observers are able to track targets despite the presence of inconsistent (outlier) measurements (Watamaniuk et al., 1994; Watamaniuk & McKee, 1995). Various techniques, known collectively as *data association* (Bar-Shalom & Fortmann, 1988), can be applied to reduce these distortions by using an additional stage that decides whether to reject or accept the new measurements. From the Bayesian perspective, this extra stage is unnecessary; robustness can be ensured by correct choices of the measurement and prior probability models. More specifically, we specify a measurement model that is robust against outliers.

2.2 Prediction and Estimation Equations for a Flow Field. We intend to estimate the motion flow field, defined over the two-dimensional image array, at each time step. We describe the flow field as $\{\vec{v}(\vec{x}, t)\}$, where the parentheses $\{\cdot\}$ denote variations over each possible spatial position $\vec{x}$ in the image array. The prior model for the motion field and the likelihood for the velocity measurements are described by the probability distribution functions $P(\{\vec{v}(\vec{x}, t)\}|\{\vec{v}(\vec{x}, t-\delta)\})$ and $P(\{\vec{\phi}(\vec{x}, t)\}|\{\vec{v}(\vec{x}, t)\})$ (see section 3 for details), where δ is time increment and $\{\vec{\phi}(\vec{x}, t)\} = \{\phi_1(\vec{x}, t), \phi_2(\vec{x}, t), \ldots, \phi_M(\vec{x}, t)\}$ represents the response of the local motion measurement units over the image array. Our theory requires that the local velocity measurements are normalized so that $\sum_{i=1}^{M} \phi_i(\vec{x}, t) = 1$ for all $\vec{x}$ and at all times t (see section 3 for details). We let $\Phi(t) = (\{\vec{\phi}(\vec{x}, t)\}, \{\vec{\phi}(\vec{x}, t-\delta)\}, \ldots)$ be the set of all measurements up to and including time t, and $P(\{\vec{v}(\vec{x}, t-\delta)\}|\Phi(t-\delta))$ be the system's estimated probability distribution of the velocity field at time $t-\delta$. Using equations 2.1 and 2.2 described above, we get

$$P(\{\vec{v}(\vec{x}, t)\}|\Phi(t)) = \frac{P(\{\vec{\phi}(\vec{x}, t)\}|\{\vec{v}(\vec{x}, t)\})P(\{\vec{v}(\vec{x}, t)\}|\Phi(t-\delta))}{P(\{\vec{\phi}(\vec{x}, t)\}|\Phi(t-\delta))}, \tag{2.3}$$

for the estimation, and

$$P(\{\vec{v}(\vec{x},t)\}|\Phi(t-\delta)) = \int P(\{\vec{v}(\vec{x},t)\}|\{\vec{v}(\vec{x},t-\delta)\})P(\{\vec{v}(\vec{x},t-\delta)\}|\Phi(t-\delta))d[\{\vec{v}(\vec{x},t-\delta)\}], \quad (2.4)$$

for the prediction.

2.3 The Marginalization Approximation. For simplicity and computational convenience, we assume that the prior and the measurement distributions are chosen to be factorizable in spatial position, so that the probabilities at one spatial point are independent of those at another. This restriction prevents us from including spatial coherence effects which are known to be important aspects of motion perception (see, for example, Yuille & Grzywacz, 1988). For the types of stimuli we are considering, these spatial effects are of little importance and can be ignored.

In mathematical terms, this spatial-factorization assumption means that we can factorize the prior P_p and the likelihood P_l so that

$$P_p(\{\vec{v}(\vec{x},t)\}|\{\vec{v}(\vec{x}',t-\delta)\}) = \prod_{\vec{x}} p_p(\vec{v}(\vec{x},t)|\{\vec{v}(\vec{x}',t-\delta)\})$$
$$P_l(\{\vec{\phi}(\vec{x},t)\}|\{\vec{v}(\vec{x},t)\}) = \prod_{\vec{x}} p_l(\{\vec{\phi}(\vec{x},t)|\vec{v}(\vec{x},t)\}). \quad (2.5)$$

A further restriction is to modify equation 2.4 so that it updates the marginal distributions at each point independently. This again reduces spatial coherence because it decouples the estimates of velocity at each point in space. Once again, we argue that this approximation is valid for the class of stimuli we are considering. This approximation will decrease computation time by allowing us to update the predictions of the velocity distributions at each point independently. The assumption that the measurement model is spatially factorizable means that the estimation stage, equation 2.3, can also be performed independently.

This gives update rules for prediction P_{pred}:

$$P_{pred}(\vec{v}(\vec{x},t)|\Phi(t-\delta)) = \int [d\vec{v}(\vec{x}',t-\delta)]P_p(\vec{v}^T(\vec{x},t)|\{\vec{v}(\vec{x}',t-\delta)\})$$
$$P_e(\{\vec{v}(\vec{x}',t-\delta)\}|\Phi(t-\delta)), \quad (2.6)$$

and for estimation P_e:

$$P_e(\vec{v}(\vec{x},t)|\Phi(t)) = \frac{P_l(\vec{\phi}(\vec{x},t)|\vec{v}(\vec{x},t))P_{pred}(\vec{v}(\vec{x},t)|\Phi(t-\delta))}{P(\vec{\phi}(\vec{x},t)|\Phi(t-\delta))}. \quad (2.7)$$

In what follows we will consider a specific likelihood and prior function that allow us to simplify these equations and define probability distributions

on the velocities at each point. This will lead to a formulation for motion prediction and estimation where computation at each spatial position can be performed in parallel.

3 The Model

We now describe a specific model for the likelihood and prior functions that, as we will show, can account for many psychophysical experiments involving temporal motion coherence. Based on this specific prior function, we then derive a formulation for motion prediction in the continuous time domain.

3.1 Likelihood Function. The likelihood function gives a probabilistic interpretation of the measurements given a specific motion. It is probabilistic because the measurement is always corrupted by noise at the input stage. (In many vision theories, the likelihood function depends on the image formation process and involves physical constraints, such as the geometry and surface reflectance. Because we are considering psychophysical stimuli, consisting of moving dots, we can ignore such complications.)

To determine our likelihood function, we must first specify the input stage. Ideally, this would involve modeling the behaviors of the cortical cells' sensing natural image motion, but this would be too complex to be practical. Instead, we use a simplified model of a bank of receptive fields tuned to various velocities $\vec{v}_i(\vec{x}, t)$, $i = 1 \cdots M$, and positioned at each image pixel. These cells have observation activities $\{\phi_i(\vec{x}, t)\}$ that are intended to represent the output of a neuronally plausible motion model such as Grzywacz and Yuille (1990). (In our implementation, these simple model cells receive contributions from the motion of dots falling within a local neighborhood, typically the four nearest neighbors. Intuitively, the closer the dot is to the center of the receptive field and the closer its velocity is to the preferred velocity of the cell, then the larger the response. The spatial profile and velocity tuning curve of these receptive fields are described by gaussian functions whose covariance matrices $\Sigma_{m:x}$ and $\Sigma_{m:v}$ depend on the direction of $\vec{v}(\vec{x}, t)$, and are specified in terms of their longitudinal and transverse components $\sigma^2_{m:x,L}$, $\sigma^2_{m:x,T}$ and $\sigma^2_{m:v,L}$, $\sigma^2_{m:v,T}$. For more details, see the appendix.)

The likelihood function specifies the probability of the receptive field responses conditioned on a "true" external motion field. Ideally this should correspond exactly to the way the motion measurements are generated. We make a simplification, however, by assuming that the measurements depend on only the velocity field at that specific position. This simplifies the mathematics at the risk of making the system slightly sensitive to motion blur (i.e., the measurement stage allows for several dots to influence the local measurement—provided the dots lie within the receptive field—but the likelihood function assumes that only a single dot causes the measurement.)

We can therefore write $P_l(\{\vec{\phi}(\vec{x},t)\}|\{\vec{v}(\vec{x},t)\}) = \prod_x P_l(\vec{\phi}(\vec{x},t)|\vec{v}(\vec{x},t))$, where $\{\vec{v}(\vec{x},t)\}$ is the external velocity field. We now specify

$$P_l(\vec{\phi}(\vec{x},t)|\vec{v}(\vec{x},t)) = \frac{\Psi_l(\phi_1(\vec{x},t),\phi_2(\vec{x},t),\ldots,\phi_M(\vec{x},t),\vec{v}(\vec{x},t))}{\int\cdots\int P_J(\xi_1,\xi_2,\ldots,\xi_M,\vec{v}_j(\vec{x},t))d\xi_1 d\xi_2,\ldots,d\xi_M}, \tag{3.1}$$

where P_J is the joint probability distribution and we set Ψ_l to be

$$\Psi_l(\phi_1(\vec{x},t),\phi_2(\vec{x},t),\ldots,\phi_M(\vec{x},t),\vec{v}(\vec{x},t)) = \vec{\phi}(\vec{x},t)\cdot\vec{f}(\vec{v}(\vec{x},t)), \tag{3.2}$$

which is attractive (see Yuille, Burgi, & Grzywacz, 1998) because it leads to a simple linear update rule, and with tuning curves f given by:

$$f_i(\vec{v}(\vec{x},t)) = \frac{e^{-(1/2)(\vec{v}_i-\vec{v})^T\Sigma^{-1}(\vec{v}_i-\vec{v})}}{\sum_{j=1}^{M} e^{-(1/2)(\vec{v}_j-\vec{v})^T\Sigma^{-1}(\vec{v}_j-\vec{v})}}, \tag{3.3}$$

where Σ is the covariance matrix that depends on the direction of $\vec{v}(\vec{x},t)$ and is specified in terms of its longitudinal and transverse components $\sigma^2_{l:v,L}$ and $\sigma^2_{l:v,T}$). The experimental data (Anstis & Ramachandran, 1987; Gottsdanker, 1956; Werkhoven et al., 1992) suggest that temporal integration occurs for velocity direction rather than for speed. This is built into our model by choosing the covariances so that the variance is bigger in the direction of motion than in the perpendicular direction (i.e., the velocity component perpendicular to the motion has mean zero and very small variance so the direction of motion is fairly accurate, but the variance of the velocity component along the direction of motion is bigger, which means that the estimation of the speed is not accurate). We have assumed that the response of the measurement device is instantaneous. It would be possible to adapt the likelihood function to allow for a time lag, but we have not pursued this. Such a model might be needed to account for motion blurring.

3.2 Prior Function. We now turn to the prior function for position and velocity. For a dot moving at approximately constant velocity, the state transitions for position and velocity are given respectively by $\vec{x}(t+\delta) = \vec{x}(t) + \delta\vec{v}(t) + \vec{\omega}_x(t)$, and $\vec{v}(t+\delta) = \vec{v}(t) + \vec{\omega}_v(t)$, where $\vec{\omega}_x(t)$ and $\vec{\omega}_v(t)$ are random variables representing uncertainty in the position and velocity. Extending this model to apply to the flow of dots in an image requires a conditional probability distribution $P_p(\vec{v}(\vec{x},t)|\{\vec{v}_j(\vec{x}_i,t-\delta)\})$, where the set of spatial positions and the set of allowed velocities are discretized with the spatial positions set to be $\{\vec{x}_i\colon i=1,\ldots,N\}$, and the velocities at $\vec{x}_i$ to $\{\vec{v}_j\colon j=1,\ldots,M\}$. We are assuming that $P_p(\{\vec{v}(\vec{x}_i,t)\}|.) = \prod_i P_p(\vec{v}(\vec{x}_i,t)|.)$ so that the velocities at each point are predicted independent of each other. We also assume that the prior is built out of two components: (1) the probability

$p(\vec{x}|\vec{x}_i, \vec{v}_j(\vec{x}_i, t-\delta))$ that a dot at position $\vec{x}_i$ with velocity $\vec{v}_j$ at time $t-\delta$ will be present at $\vec{x}$ at time t and (2) the probability $P(\vec{v}(\vec{x}, t)|\vec{v}_j(\vec{x}_i, t-\delta))$ that it will have velocity $\vec{v}(\vec{x}, t)$. These components are combined to give:

$$P_p(\vec{v}(\vec{x}, t)|\{\vec{v}_j(\vec{x}_i, t-\delta)\}) = \frac{1}{K(\vec{x}, t)} \sum_i \sum_j P(\vec{v}(\vec{x}, t)|\vec{v}_j(\vec{x}_i, t-\delta))\, p(\vec{x}|\vec{x}_i, \vec{v}_j(\vec{x}_i, t-\delta)), \quad (3.4)$$

where $K(\vec{x}, t)$ is a normalization factor chosen to ensure that $P_p(\vec{v}(\vec{x}, t)|\{\vec{v}_j(\vec{x}_i, t-\delta)\})$ is normalized. The two conditional probability distribution functions in equation 3.4 are assumed to be normally distributed, and thus,

$$p(\vec{x}|\vec{x}_i, \vec{v}_j(\vec{x}_i, t-\delta)) \sim e^{-(1/2)(\vec{x}-\vec{x}_i-\delta\vec{v}_j(\vec{x}_i,t-\delta))^T \Sigma_x^{-1} (\vec{x}-\vec{x}_i-\delta\vec{v}_j(\vec{x}_i,t-\delta))} \quad (3.5)$$

$$P(\vec{v}(\vec{x}, t)|\vec{v}_j(\vec{x}_i, t-\delta)) \sim e^{-(1/2)(\vec{v}(\vec{x},t)-\vec{v}_j(\vec{x}_i,t-\delta))^T \Sigma_v^{-1} (\vec{v}(\vec{x},t)-\vec{v}_j(\vec{x}_i,t-\delta))}. \quad (3.6)$$

These functions express the belief that the dots are, on average, moving along a straight trajectory (see equation 3.5) with constant velocity (see equation 3.6). The covariances Σ_x and Σ_v, which quantify the statistical deviations from the model, are defined in a coordinate system based on the velocity vector $\vec{v}_j$. These matrices are diagonal in this coordinate system and are expressed in terms of their longitudinal and transverse components $\sigma_{x,L}^2$, $\sigma_{x,T}^2$, $\sigma_{v,L}^2$, $\sigma_{v,T}^2$. (This model predicts future positions and velocities indpendently; it does not take into account the possibility that if the speed increases during the motion, then the dot will travel further. This is an acceptable, and standard, approximation provided that δ is small enough. It becomes exact in the limit as $\delta \mapsto 0$.)

3.3 Continuous Prediction. Computer simulations of equation 2.6 (even for the simple case of moving dots) require large-kernel convolutions, which require excessive computational time. Instead, we reexpress equation 2.6 in the continuous time domain. (Such a reexpression is not always possible and depends on the specific choices of probability distributions. Note that a similar approach has been applied for the computation of stochastic completion fields by Williams and Jacobs (1997a, 1997b).) As shown in the appendix for Gaussian priors, the evolution of $P(\vec{v}(\vec{x}, t)|\Phi(t-\delta))$ for $\delta \to 0$ satisfies a variant of a partial differential equation, known as Kolmogorov's forward equation or Fokker-Planck equation. Our equation is a nonstandard variant because, unlike standard Kolmogorov theory, our theory involves probability distributions at every point in space that interact over time. It is given by:

$$\frac{\partial}{\partial t} P(\vec{v}(\vec{x}, t)) = \frac{1}{2} \left\{ \frac{\partial^T}{\partial \vec{v}} \Sigma_{\vec{v}} \frac{\partial}{\partial \vec{v}} \right\} P(\vec{v}(\vec{x}, t)) + \frac{1}{2} \left\{ \frac{\partial^T}{\partial \vec{x}} \Sigma_{\vec{x}} \frac{\partial}{\partial \vec{x}} \right\} P(\vec{v}(\vec{x}, t))$$

$$-\vec{v}\cdot\frac{\partial}{\partial\vec{x}}P(\vec{v}(\vec{x},t))+\frac{1}{|V|}\frac{\partial}{\partial\vec{x}}\int\vec{v}P(\vec{v}(\vec{x},t))d\vec{v}. \tag{3.7}$$

where $|V|$ is the volume of velocity space $\int d\vec{v}$, and $\{\frac{\partial^T}{\partial\vec{x}}\Sigma_{\vec{x}}\frac{\partial}{\partial\vec{x}}\}$ and $\{\frac{\partial^T}{\partial\vec{v}}\Sigma_{\vec{v}}\frac{\partial}{\partial\vec{v}}\}$ are scalar differential operators (for isotropic diffusions, these scalar operators are Laplacian operators). The terms on the right-hand side of this equation stand for velocity diffusion (first term), spatial diffusion (second term) with deterministic drift (third term), and normalization (fourth term). This equation is local and describes the spatiotemporal evolution of the probability distributions $P(\vec{v}(\vec{x},t))$.

4 The Implementation

We now address the implementation of the update rule for motion estimation (see equations 2.7). At each spatial position, two separate cells' populations are assumed: one for coding the likelihood function and another for coding motion estimation. If we were to choose the large-kernel convolution for motion prediction (see equation 2.6), then the network implementation would reduce to two positive linear networks multiplying each other followed by a normalization, illustrated in Figure 1 (see Yuille et al., 1998). But a problem with implementing this network on serial computers is that the range of the connections between motion estimation cells depends on the magnitude of $\vec{v}_j$ (that is, we would need long-range connections for high speed tunings). Such long-range connections occur in biological systems, like the visual cortex, but are not well suited to very large-scale integrated or serial computer implementations. Alternatively, using our differential equation for predicting motion (see equation 3.7) involves only local connections. We now focus on this new implementation (see Figure 2).

Kolmogorov's equation can be solved on an arbitrary lattice using a Taylor series expansion. The spatial term, which contains the diffusion and drift terms, can be written so that the partial derivatives at lattice points are functions of the values of the closest neighbor points. For the sake of numerical accuracy, we have chosen a spatial mesh system where the points are arranged hexagonally. Furthermore, we found it convenient to express the velocity vectors in polar coordinates where their norm s and angle θ can be represented on a rectangular mesh. The change of variable is given by $\tilde{p}(s,\theta)=sp(\vec{v})$. In polar coordinates, the differential operator $\frac{\partial^T}{\partial\vec{v}}\Sigma_{\vec{v}}\frac{\partial}{\partial\vec{v}}$ becomes

$$L_v[\tilde{p}]=\frac{1}{2}\sigma_{v,L}^2\frac{\partial^2\tilde{p}}{\partial s^2}-\left(\sigma_{v,L}^2-\frac{1}{2}\sigma_{v,T}^2\right)\frac{\partial}{\partial s}\left(\frac{1}{s}\tilde{p}\right)+\frac{1}{2}\sigma_{v,T}^2\frac{\partial^2\tilde{p}}{\partial\theta^2}\frac{1}{s^2}, \tag{4.1}$$

where $\sigma_{v,L}^2$ and $\sigma_{v,T}^2$ are the variances in the longitudinal and transverse directions, respectively.

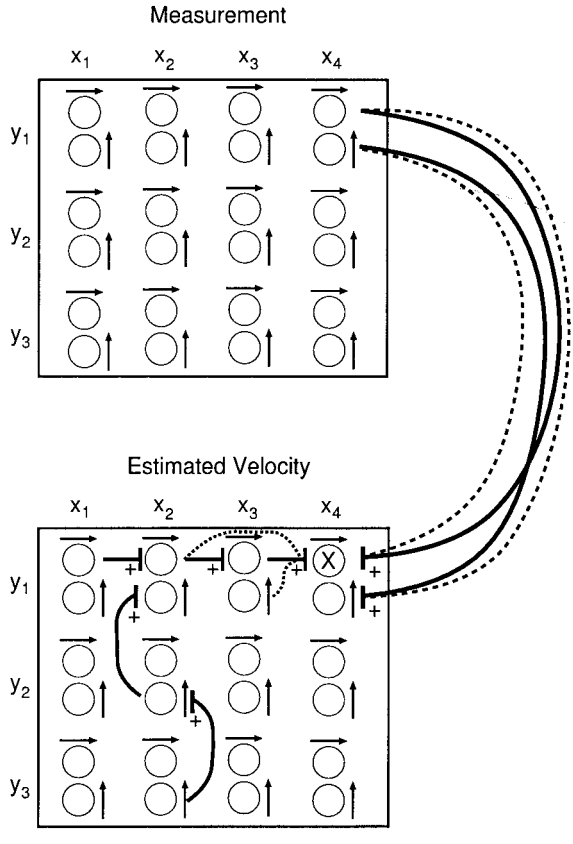

Figure 1: Network for motion estimation. The network consists of two interacting layers. The top square shows the observation layer consisting of cells organized in columns on the (x, y) lattice. At each of the 12 spatial positions, there is a velocity column, which we display by two cells, shown as circles, with the adjacent arrows indicating their velocity tunings (here either horizontal or vertical). The lower square represents the estimation layer, which also consists of cells organized as columns on the (x, y) lattice. In the measurement stage, the observation cells influence the estimation cells by excitatory (indicated by the plus sign) connections and multiplicative interactions (indicated by the cross sign). The excitation is higher between cells with similar velocity tuning, which we indicate by strong (solid lines) or weak (dashed lines) connections. In the prediction stage, cells within the estimation layer excite each other (again with the strength of the excitation being largest for cells with similar velocity tuning). Inhibitory connections within the columns are used to ensure normalization.

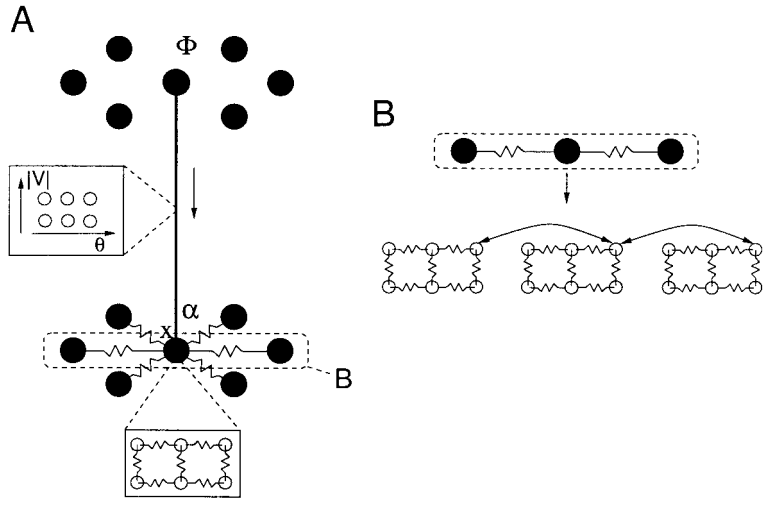

Figure 2: Network for motion estimation using continuous prediction. (A) Continuous motion prediction (lower part of the network) is accomplished through nearest-neighbor spatial interactions—in this case, on a hexagonal lattice. The velocity distribution is represented at each spatial position by a set of velocity cells (small circles) organized according to a polar representation (small panel). Observation cells in the measurement layer (upper part) influence the estimation cells by multiplicative interactions (indicated by $\times$) to yield the motion estimation α. (B) Interactions between two populations of velocity cells involve cells tuned to the same velocities; interactions within a population involve cells tuned to different velocities.

We let the activities of the M motion estimation cells at time t be represented by a vector $\vec{\alpha}(\vec{x}, t) = (\alpha_1(\vec{x}, t), \ldots, \alpha_M(\vec{x}, t))$. The iterative method for determining such activities involves a four-step scheme. The first three steps are concerned with solving Kolmogorov's equation. For a cell at position $\vec{x} = (x, y)$ and of velocity tuning $\vec{v}_\mu = (s, \theta)$, the three steps are:

$$\begin{aligned}
\alpha^{t+1/4}_{x,y,s,\theta} &= \alpha^{t}_{x,y,s,\theta} + \Delta t \sum_{\iota,} A_{\iota,}(s)\alpha^{t}_{x,y,s+\iota\Delta s,\theta+\Delta\theta} \\
\alpha^{t+1/2}_{x,y,s,\theta} &= \alpha^{t+1/4}_{x,y,s,\theta} + \Delta s \sum_{\iota,\chi} B_{\iota,\chi}(s,\theta)\alpha^{t+1/4}_{x+\iota,y+\chi,s,\theta} \\
\alpha^{t+3/4}_{x,y,s,\theta} &= \alpha^{t+1/2}_{x,y,s,\theta} - \frac{\Delta t}{M} \sum_{s',\theta'} \sum_{\iota,\chi} B_{\iota,\chi}(s',\theta')\alpha^{t+1/2}_{x+\iota,y+\chi,s',\theta'}
\end{aligned} \tag{4.2}$$

where the coefficients $A_{\iota,}$ and $B_{\iota,\chi}$ are functions of s, θ, and the covariance matrices. The constants Δs, $\Delta\theta$ represent the quantization factors for s and θ. The superscripts $t + 1/4$, $t + 1/2$, $t + 3/4$ indicate the order in which these

three steps proceed (a single time step for the complete system is broken down into four substeps for implementational convenience).

The first step is to evaluate the velocity differential operator and involves four neighbor cells. The second step calculates the spatial differential operator and involves six neighbor points (on the hexagonal spatial lattice). Periodic boundary conditions are assumed in space and velocity. The third step performs the normalization. To guarantee stability of the whole iterative method, the time step Δt has been determined using von Neumann analysis (see Courant & Hilbert, 1953), which considers the independent solutions, or eigenmodes, of finite-difference equations. Typically the stability conditions involve ratios between the time step and the quantization scales of space and velocity.

If we are performing prediction without measurement, then the fourth step—determining the activity of the motion estimation cells—simply reduces to

$$\alpha_{x,y,s,\theta}^{t+1} = \alpha_{x,y,s,\theta}^{t+3/4}. \tag{4.3}$$

If there are measurements, then we apply the motion estimation equation, 2.7, which consists of multiplying the motion prediction with the likelihood function (as defined by equations 2.8 and 2.9). Therefore, the complete update rule is

$$\alpha_{x,y,s,\theta}^{t+1} = \frac{1}{N(\vec{x}, t+1)} \alpha_{x,y,s,\theta}^{t+3/4} \cdot \sum_j \phi_j(\vec{x}, t) \cdot f_j(\vec{v}_\mu(\vec{x}, t)), \tag{4.4}$$

where $N(\vec{x}, t+1)$ is a normalization constant to impose $\sum_{\mu=1}^{M} \alpha_\mu(\vec{x}, t+1) = 1, \ \forall \vec{x}, t$ so that we can interpret these activities as the probabilities of the true velocities.

5 Methods

Our model has two extreme forms of behavior depending on how well the input data agree with the model predictions. If the data are generally in agreement with the model's predictions, then it behaves like a standard Kalman filter (i.e., with gaussian distributions) and integrates out the noise. At the other extreme, if the data are extremely different from the prediction, then the model treats the data as outliers or distractors and ignores them. We are mainly concerned here with situations where the model rejects data. The precise situation where noise and distractors start getting confused is very interesting, but we do not address it here.

We first checked the ability of our model to integrate out weak noisy stimuli for tracking a single dot moving in an empty background. For this situation, temporal grouping (data association) is straightforward, and so

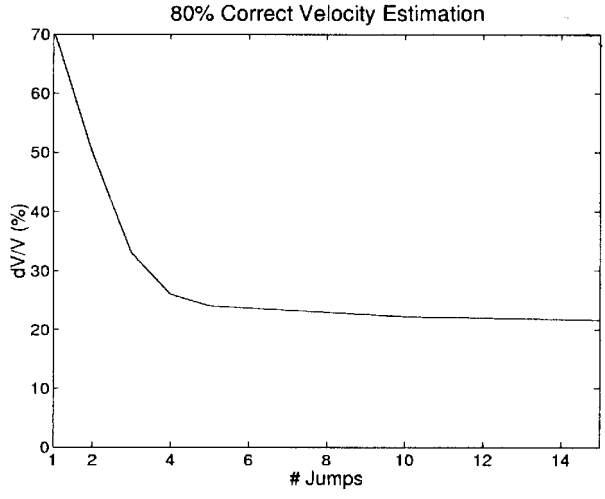

Figure 3: Speed discrimination for two moving dots. This graph shows the improved ability of our theory to estimate the relative speed of two dots (one of which is moving at 2 jumps per second) as a function of the number of jumps. The level of 80% correct discrimination is reached for diminishing speed differences (measured by dV/V) when the number of jumps increases, consistent with psychophysics (McKee et al., 1986).

standard Kalman filters can be used. To evaluate the model, we ran a set of trials that plotted the relative estimation of two velocities as a function of the number of jumps. The graph we obtained (see Figure 3) is very similar to those reported in the psychophysical literature (McKee et al., 1986).

Henceforth, we assume that the model is able to integrate out noisy input. For the remaining experiments, we assumed that the local velocity measurements were "noisy" in the sense that the measurement units specified a probability of velocity measurements rather than a single velocity. However, the "probability of velocities" itself was not noisy.

We next tested our model on three psychophysical motion phenomena. First, we considered the improved accuracy of velocity estimation of a single dot over time (Snowden & Braddick, 1989; Watamaniuk, Sekeuler, & Williams, 1989; Welch, Macleod, & McKee, 1997; Ascher, 1998). Then we examined the predictions for the motion occlusion experiments (Watamaniuk & McKee, 1995). Finally we investigated the motion outlier experiments (Watamaniuk et al., 1994).

For all simulations, we set the parameters as follows. All dots were moving at a speed of 6 jumps per second (jps), where one jump corresponds to the distance separating two neighboring pixels in the horizontal direction. The longitudinal and transverse components of the covariances matrices were

$\sigma_{m:x,L} = 0.8$ jump, $\sigma_{m:x,T} = 0.4$ jump, $\sigma_{m:v,L} = 3.2$ jps, and $\sigma_{m:v,T} = 2.6$ jps for the measurement, $\sigma_{l:v,L} = 2.2$ jps, and $\sigma_{l:v,T} = 1.1$ jps for the likelihood function, and $\sigma_{x,L} = 0.6$ jump, $\sigma_{x,T} = 0.3$ jump, $\sigma_{v,L} = 0.8$ jps, and $\sigma_{v,T} = 0.4$ jps for the prior function. These parameters were chosen by experimenting with the computer implementation, but the results are relatively insensitive to their precise values; detailed fine tuning was not required.

Except for one of the occluder experiments (the occluder defined by distractor motion), where we used 18 velocity channels (six equidistant directions and three speeds), we used 30 velocity channels positioned at each spatial position, that is, six equidistant directions (thus $\Delta\theta = 60$ degrees, starting at the origin) and five speeds ($\Delta r = 2$ jps, with the slowest speed channel tuned to 2 jps). To guarantee stability of the numerical method, we set the time increment to $\Delta t = 0.6$ ms. Initial conditions were uniformly distributed density functions. There were 32×32 spatial positions.

6 Results

6.1 Velocity Estimation in Time. For this experiment, the stimulus was a single dot moving in an empty background. To evaluate how the velocity estimation evolves in time, we measured two numbers. The first number was the *sharpness* of the velocity probability at each position, which we define to be the negative of the entropy of the distribution (the entropy of the distribution is given by $-\sum_{j=1}^{M} \alpha_j(\vec{x}, t) \log \alpha_j(\vec{x}, t)$) plus the maximum of the entropy (to ensure that the sharpness is nonnegative). Observe that the sharpness is the Kullback-Leibler divergence $D(\alpha \| U)$ between the velocity distribution $\alpha_j(\vec{x}, t)$ and the uniform distribution $U_j = 1/M, \ \forall\, j$ (more precisely, $D(\alpha \| U) = \sum_{j=1}^{M} \alpha_j(\vec{x}, t) \log\{\alpha_j(\vec{x}, t)/U_j\}$.) As is well known (Cover & Thomas, 1991) the Kullback-Leibler divergence is positive semidefinite, taking its minimum value of 0 when $\alpha_j = U_j, \ \forall j$ and increases the more α_j differs from the uniform distribution U_j (i.e., the sharper the distribution α_j becomes). Hence the higher this sharpness is, the more precise the velocity estimate is. Moreover, the sharpness will be lowest in positions where there are no dots moving (i.e., for which $\phi_j(\vec{x}, t) = 1/M, j = 1, \ldots, M$). The target, a coherently moving dot, should have a relatively high sharpness response surrounded by low sharpness responses in neighboring positions. The second number is the *confidence factor* $P(\vec{\phi}(\vec{x}, t) | \Phi(t - \delta))$ in equation 2.7. This measures how probable the model judges the new input data. The higher this number is, the more the new measurement is consistent with the prediction (and hence the more likely that the measurement is due to coherent motion).

The results of motion estimation for this experimental stimulus are shown in Figure 4. This figure shows an initial fast decrease in the entropy of the velocity distribution followed by a slow decrease (i.e., an increase in sharpness of the density function). Also shown is the increase in the confidence

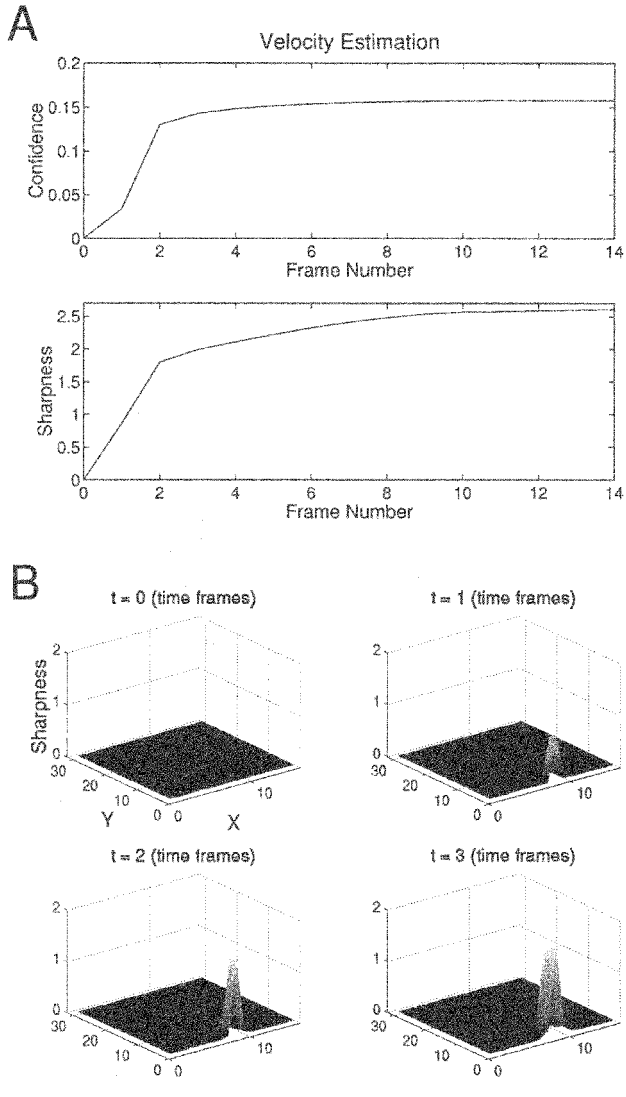

Figure 4: Single dot experiment. A single dot is moving with constant velocity. (A) Increasing accuracy of the velocity estimate with time is visible by an increase in the confidence factor and the sharpness of the density function. Observe how both the confidence and sharpness increase rapidly at first but then quickly start to flatten out. (Our model outputs a data point at each time frame, and our plotting package joins them together by straight-line segments.) (B) Sharpness of the density function shown for each spatial position at successive time frames. The system converges to a single peak in the velocity estimation (though we do not show this here).

factor of the target velocity estimation at each new iteration. These two effects indicate the increased accuracy of the velocity estimation with each new iteration, consistent with the psychophysical literature (Snowden & Braddick, 1989; Watamaniuk el al., 1989; Welch et al., 1997; Ascher, 1998). Note that for coherent motion, the sharpness of the model increases rapidly at first and then appears to slow down. Our results suggest that the model reaches an asymptotic limit, although they do not completely rule out a continual gradual increase. We argue that asymptotic behavior is more likely as both the predictions and the measurements have inherent noise that can never be eliminated. Moreover, it can be proven that the standard (gaussian) Kalman model will converge to an asymptotic limit depending on the variances of the prior and likelihood functions (we verified this by computer simulations). It seems also that human observers reach a similar asymptotic limit, and their accuracy does not become infinitely good with an increasing number of input stimuli. In addition, we plot the sharpness of the velocity estimates as a function of spatial position in Figure 4B (each column of velocity-tuned cells has a sharpness associated with its velocity estimate).

We also tested the estimation of velocity for a dot moving on a circular trajectory, and our model can successfully estimate the dot's speed, although the sharpness and confidence are not as high as they are for straight motion (recall that the prior prefers straight trajectories). This again seems consistent with the psychophysics (Watamaniuk et al., 1994; Verghese et al., 1999).

6.2 Single Dot with Occluders. Next, we explored what would happen if the target dot went behind an occluding region where no input measurements could be made (we call this a *black occluder*). The results were the same as the previous case until the target dot reached the occluder. In the occluded region, the motion model continued to propagate the target dot, but, lacking supporting observations, the probability distribution started to diffuse in space and in velocity. This dual diffusion can be seen in Figure 5, where the sharpness of velocity estimation is shown after the dot entered the occluding region. Observe the decrease in sharpness of the density function, indicating a degradation of velocity estimation, when the target is behind the occluder. However, the model still has enough "momentum" to propagate its estimate of the target dot's position even though no measurements are available (this will break down if the occluder gets too large). This is consistent with the findings by Watamaniuk and McKee (1995), who showed that observers had a higher-than-average chance of detecting the target when it emerges from the occluder into a noisy background.

We then tested our model with *motion occluders*—occluders defined by motion flows as described in Watamaniuk and McKee (1995) (see Figure 6A). We use the same measures as for the single isolated dot. We also plot the cells' activities as a function of the direction tuning for the cells tuned to the optimal speed (see Figure 6B). The plot indicates that the cells can signal nonzero probabilities for several motions at the same time. After entering the

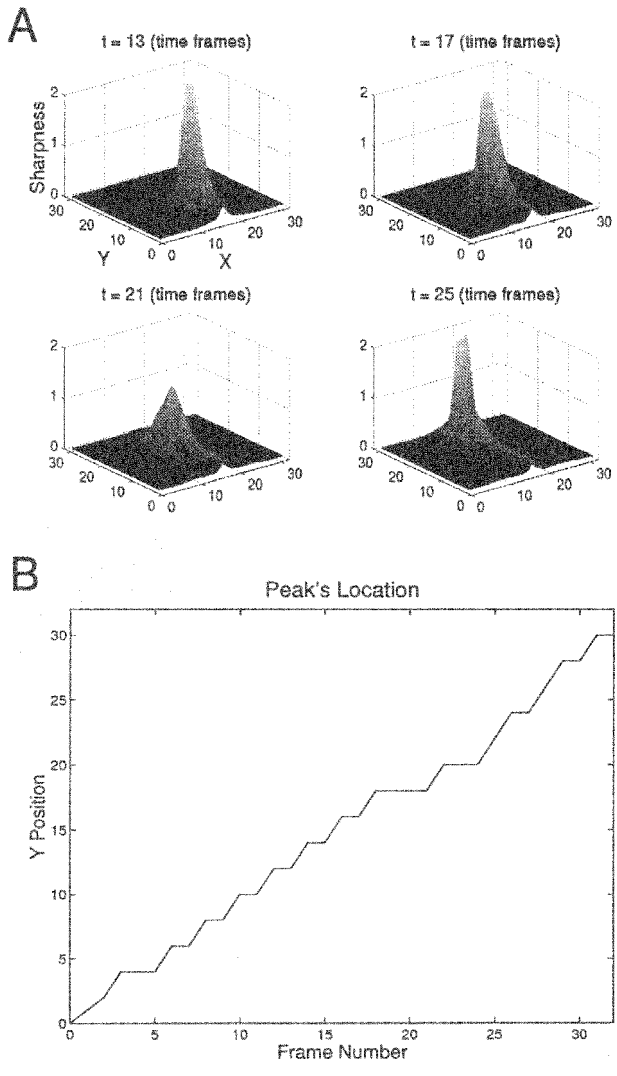

Figure 5: Single dot experiment with Black occluder. (A) This representation shows how the spatial blur and sharpness of the velocity distribution change with time (measured in time frames) when a single dot moving along the y-axis gets occluded and reappears. The occluding area is from $y = 15$ up to $y = 22$. (B) The y-location of the peak of velocity distribution as a function of time. Motion inertia (momentum) keeps the peak moving during the occlusion, albeit with a tendency to slow down (as visible by the wider plateau around time frame 20).

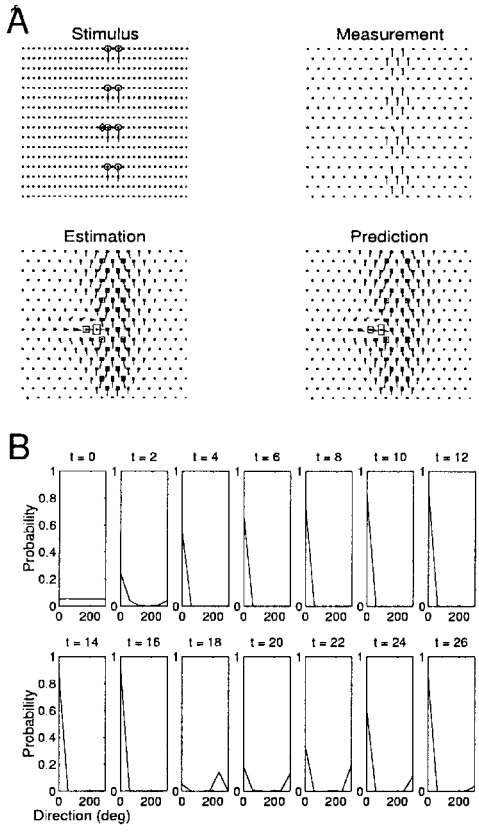

Figure 6: Single dot experiment with occluders defined by distractor motion. (A) The occluder is defined by the vertical motion of distractor dots, visible as circles in the left-top frame with arrows indicating velocity (the dots' speed is 6 jumps per second and is represented by the length of the lines). The target motion, shown at time frame 15 and visible as a diamond, is moving horizontally from left to right. The height and width of the rectangles at each point indicate the confidence and sharpness, respectively (a small rectangle indicates low confidence and sharpness). The lattice, marked with small dots, is hexagonal for the measurement, estimation, and prediction stages. (B) The probability distribution of the velocity cells tuned to 6 jumps per second plotted as a function of directional tuning at different time steps. The motion-inertia effect of the target motion on the distractors is visible at time frame 18 as the target dot enters the occluder and two peaks start developing in the probability distribution. The bigger the occluder, the more the peak induced by the motion of the distractor dots starts dominating. But as the dot reemerges from the occluder, it rapidly becomes sharp again, as visible at time frame 20.

occluding region, the peak corresponding to the target motion gets smaller than the peak due to the occluders. However, the peak of the target remains, and so the target peak can increase rapidly as the target exits from the occluders. Our model predicts that it is easier to deal with black occluders than with motion occluders, which is not consistent with the psychophysics that shows a rough similarity of effects for both types of occluders. We offer two possible explanations for this inconsistency. First, the model's directionally selective units and/or prediction-propagation mechanism have too broad a directional tuning. Narrowing it may lead to weaker interactions between the motion occluder and the target. Second, in contrast with the black occluders, the motion occluders may group together as a surface, which would help explain why the visual system may be more tolerant of motion occluders. (Such a heightened tolerance might compensate for the initial putative advantage of the black occluders.) Kanizsa (1979) demonstrates several perceptual phenomena, which appear to require this type of explanation. In motion, similar effects have been modeled by Wang and Adelson (1993) in their work on layered motion. A limitation of our current model is that it does not include such effects.

6.3 Outlier Detection. In the outlier detection experiments (Watamaniuk et al., 1994; Verghese et al., 1999), the target dot is undergoing regular motion, but it is surrounded by distractor dots, which are undergoing random Brownian motion. To show the velocity estimation at each position, we plot the response of our network using a rectangle to display the two properties of sharpness and confidence. The width of the rectangle gives the sharpness of the set of cells at that point, and the height gives the confidence factor. The arrow shows the mean of the estimated velocity. It can be seen in Figure 7 that the target dot's signal rapidly reaches large sharpness and confidence by comparison to the distractor dots, which are not moving coherently enough to gain confidence or sharpness. The sharpness of the target dot's signal does not grow monotonically because the distractor dots sometimes interfere with the target's motion by causing distracting activity in the measurement cells.

7 Discussion

This work provides a theory for temporal coherence. The theory was formulated in terms of Bayesian estimation using motion measurements and motion prediction. By incorporating robustness in the measurement model, the system could perform temporal grouping (or data association), which enabled it to choose what data should be used to update the state estimation and what data could be ignored. We also derived a continuous form for the prediction equation, and showed that it corresponds to a variant of Kolmogorov's equations at each node. In deriving our theory, we made an assumption of spatial factorizability of the probability distributions and

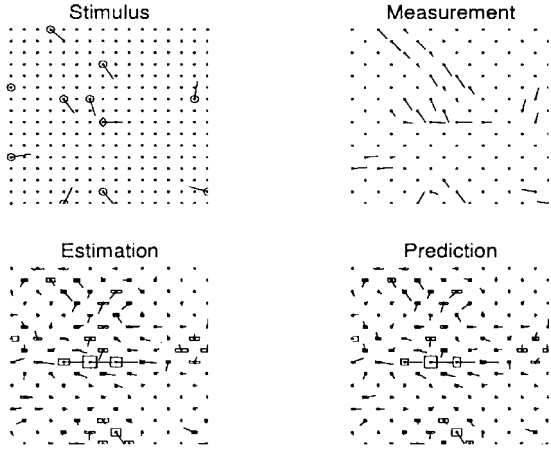

Figure 7: Outlier detection. A target dot (shown as a diamond with line pointing rightward, indicating its velocity) is halfway down the stimulus frame moving from left to right with noise-free motion. It is surrounded by distractor dots undergoing Brownian motion. After a few time step, the model gives high confidence and sharpness for the target dot. Observe that motion estimates of the distractor dots can sometimes become sharp if their motion direction is not changing too radically between two time frames (see the large rectangles at the bottom of the estimation and prediction panels.) Graphic conventions as in Figure 6.

made the approximation of updating the marginal distributions of velocity at each point. This allowed us to perform local computations and simplified our implementation. We argued that these assumptions and approximations (as embodied in our choice of prior probability) are suitable for the stimuli we are considering, but they would need to be revised to include spatial coherence effects (Yuille & Grzywacz, 1988, 1989). The prior distribution used in this article would need to be modified to incorporate these effects.

In addition, recent results by McKee and Verghese (personal communication) suggest that a more sophisticated prior may be needed. Their results seem to show that observers are better at predicting *where* there is likely to be motion rather than *what* the velocity will be. Recall that our prior probability model has two predictive components based on the current estimation of velocity at each image position. First, the velocity estimation is used to predict the positions in the image where the velocity cells would be excited. Second, we predict that the new velocity is close to the original velocity. By contrast, McKee and Verghese's results seem to show that human observers can cope with reversals of motion direction (such as a ball bouncing off a wall). This is a topic of current research, and we note that priors that al-

low for this reversal have already been developed by the computer vision tracking community (Isard & Blake, 1998).

Simulations of the model seem in qualitative agreement with a number of existing psychophysical experiments on motion estimation over time, motion occluders and motion outliers (McKee et al., 1986; Watamaniuk & McKee, 1995; Watamaniuk et al., 1994). Such a qualitative agreement is characterized by three main features: (1) the final covariance of the motion estimate, (2) the number of jumps needed to reach such a final covariance, and (3) the extent of motion inertia during an occlusion, or between two measurements. These three features are controlled as follows: the likelihood's covariance matrix determines the initial distribution, and thus the number of jumps to reach the desired covariance, while the prior's covariance matrix affects all three features by imposing an upper bound to the covariance and setting the time constants of the diffusion process between measurements.

Implementation of the model's equations led to a network that shares interesting similarities with known properties of the visual cortex. It involves a columnar structure where the columns, at each spatial point, are composed of a set of cells tuned to a range of velocities (columnar structures exist in the cortex but none has yet been found for velocities). The observation cell layer involves local connections within the columns to compute the likelihood function. The estimated velocity flow is computed in a second layer. If these layers exist, it would be natural for them to lie in cortical areas involved with motion such as MT and MST (Maunsell & Newsome, 1987; Merigan & Maunsell, 1993). The estimation layer carries out the calculation according to Kolmogorov's equation, or the discrete long-range version described in Yuille et al. (1998), and consists of spatial columns of cells that excite locally each other to predict the expected velocities in the future. The calculation requires excitatory and inhibitory synapses (i.e., no "negative synapses"). The inputs from the observation layer multiply the predictions in the estimated layer. Finally, we postulate that inhibition occurs in each column to ensure normalization of the velocity estimates at each column. This normalization could be implemented by lateral inhibition as proposed by Heeger (1992).

A difficult aspect of this network, as a biological model, is its need for multiplications (such multiplications arise naturally from our probabilistic formulation using the rules for combining probabilities). Neuronal multiplication mechanisms were argued for by Reichardt (1961) and Poggio and Reichardt (1973, 1976) on theoretical grounds. A specific biophysical model to approximate multiplication was proposed by Torre and Poggio (1978). Detailed investigations of this model, however, showed that it provided at best a rough approximation for multiplication (Grzywacz & Koch, 1987). Moreover, experiments showed that motion computations in the rabbit retina, for which the model was originally proposed, were not well approximated by multiplications (Grzywacz, Amthor, & Mistler, 1990). Recent related work (Mel, Ruderman, & Archie, 1988; Mel, 1993), though arguing

for multiplication-like processing, has not attempted to attain good approximations to multiplication. On the other hand, the complexities of neurons make it hard to rule out the existence of domains in which they perform multiplications. Overall, it seems best to be agnostic about neuronal multiplication and trust that more detailed experiments will resolve this issue. More pragmatically, it may be best to favor neuronal models that correspond to known biophysical mechanisms. The biophysics of neurons may prevent them from performing "perfect" Bayesian computations, and perhaps one should instead seek optimal models that respect these biophysical limitations.

Finally, we emphasize that simple networks of the type we propose can implement Bayesian generalizations of Kalman filters for estimation and prediction. The popularity of standard linear Kalman filters is often due to pragmatic reasons of computational efficiency. Thus, linear Kalman filters are often used even when they are inappropriate as models of the underlying processes. In contrast, the main drawback of such Bayesian generalizations is the high computational cost, although statistical sampling methods can be used successfully in some cases (see Isard & Blake, 1996). Our study shows that these computational costs may be better dealt with by using networks with local connections, and we are investigating the possibility of implementing our model on parallel architectures in the expectation that we will then be able to do Bayesian estimation of motion in real time.

Appendix

A.1 Receptive Field Tunings. More precisely, we set:

$$\phi_i(\vec{x}, t) = \frac{\psi_i(\vec{x}, t)}{\sum_{j=1}^{N} \psi_j(\vec{x}, t)},$$

where

$$\begin{aligned}\psi_i(\vec{x}, t) = A &+ \lambda_1 \sum_{\vec{x}' \in Nbh(\vec{x})} G(\vec{x}' - \vec{x} : \widehat{\vec{v}^T}(\vec{x}', t), \sigma^o_{x,t}, \sigma^o_{x,l}) \\ &\times G(\vec{v}^T(\vec{x}', t) - \vec{v}_i : \vec{v}^T(\vec{x}', t), \sigma^o_{v,t}, \sigma^o_{v,l}) \\ &+ (1 - \lambda_1) \sum_{\vec{x}' \in Nbh(\vec{x})} G(\vec{x}' - \vec{x} : \widehat{\vec{v}^T}(\vec{x}', t - \delta), \sigma^o_{x,t}, \sigma^o_{x,l}) \\ &\times G(\vec{v}^T(\vec{x}', t - \delta) - \vec{v}_i : \vec{v}^T(\vec{x}', t - \delta), \sigma^o_{v,t}, \sigma^o_{v,l}),\end{aligned}$$

and the three terms of the right-hand side of the equation are explained as follows.

The first term, A, is the rest activity of the cells. It ensures that we will have observations $\phi_i(\vec{x}, t) = 1/N$, $\forall i$ if no velocity is present. In other words, if no dot moves across the cell, then all velocities are considered to be equally likely.

The second term contains summations over true velocities (i.e., dots) within the receptive field of the cell. It contains a spatial decay term $G(\vec{x}' - \vec{x}: .,.,.)$ and a velocity tuning curve $G(\vec{v}^T(\vec{x}') - \vec{v}_i: .,.,.)$. G is a gaussian function with longitudinal and tranverse variances $\sigma_{.,l}$ $\sigma_{.,t}$ defined in terms of the direction $\hat{\vec{v}}^T(\vec{x}', t)$. In addition, the velocity tuning can depend on the speed $\left|\vec{v}^T(\vec{x}', t)\right|$. The observation model therefore assumes that the cell can measure the components of position and velocity most accurately in the direction perpendicular to the true motion of the stimulus. Intuitively, the closer the dot is to the center of the receptive field and the closer its velocity is to the preferred velocity of the cell, then the larger the response is. In realistic motions (i.e., not dots!) we can typically assume that the velocity will be constant within each cell except at motion boundaries and transparencies (though there are receptive fields at multiple scales, and so the velocity may not be constant within the large cells).

The third term is similar to the second, except that it contains memory of the true velocity at time $t - \delta$. This "memory" is due to the delayed response of the cell. $0 \leq \lambda_1 \leq 1$, so the third term is always positive.

The response of the observation cells at a point $\vec{x}$ at time t can be thought of as a local estimate of the probability of velocity. If there is no true velocity to be observed, then the response of all these cells would be uniform (i.e., $\phi_i(\vec{x}, t) = 1/N, \ \forall \ i$). If the cells receive input from a single true velocity, then the observation cell best tuned to this velocity will have the biggest response. The closer the true velocity stimulus (i.e., the dot) is to the center of the cell, then the stronger the peak. By contrast, multiple peaks can occur if there are several motions in the receptive field. This can occur for motion transparency, near-motion boundaries, and occluding motion. It can also occur in our model if there is one true motion in the receptive field at time $t - \delta$ and a different motion at time t.

A.2 Kolmogorov's Equation. In this appendix, we derive Kolmogorov's equation for motion prediction. Let us consider equation 2.6 for $\delta \to 0$. Taking this limit is tricky and requires Itô calculus (Jazwinski, 1970). Our case is also nonstandard because our theory involves probability distributions at every point in space, which interact over time. This means that we cannot simply use the standard Kolmogorov's equations, which apply to a single distribution only. Here we give a simple derivation that which uses delta functions and is appropriate when the prior model is based on a gaussian $G(.)$. The prior specified by equations 3.4 and 3.5 is written as:

$$P(\vec{v}(\vec{x}, t+\delta)|\{\vec{v}'(\vec{x}', t)\}) = \int d\vec{x}' \int d\vec{v}'(\vec{x}') G(\vec{v}(\vec{x}, t+\delta) - \vec{v}'(\vec{x}', t) : \delta\Sigma_{\vec{v}}) \times \frac{G(\vec{x} - \vec{x}' - \delta\vec{v}'(\vec{x}', t) : \delta\Sigma_{\vec{x}})}{\int d\vec{x}'' \int d\vec{v}''(\vec{x}'') G(\vec{x} - \vec{x}'' - \delta\vec{v}''(\vec{x}'', t); \delta\Sigma_{\vec{x}})}. \quad \text{(A.1)}$$

The normalization term in the denominator is used to ensure that the probabilities of the velocities integrate to one at each spatial point $\vec{x}$. In this equation, we have scaled the covariances Σ by δ. This is necessary for taking the limit as $\delta \mapsto 0$ (Jazwinski, 1970).

Between observations we can express the evolution of the prior density $P(\vec{v}(\vec{x}, t))$ as

$$\begin{aligned}\frac{\partial P(\vec{v}(\vec{x}, t))}{\partial t} &= \lim_{\delta \mapsto 0} \left\{ \frac{P(\vec{v}(\vec{x}, t+\delta) | \{\vec{v}'(\vec{x}', t)\}) - P(\vec{v}(\vec{x}, t))}{\delta} \right\} \\ &= \lim_{\delta \mapsto 0} \frac{1}{\delta} \left\{ \int d\vec{x}' \int d\vec{v}'(\vec{x}') P(\vec{v}(\vec{x}, t+\delta) | \{\vec{v}'(\vec{x}', t)\}) P(\{\vec{v}'(\vec{x}', t)\}) \right. \\ &\qquad \left. - P(\vec{v}(\vec{x}, t)) \right\}. \end{aligned} \tag{A.2}$$

We now perform a Taylor series expansion of $P(\vec{v}(\vec{x}, t+\delta) | \{\vec{v}'(\vec{x}', t)\})$ in powers of δ keeping the zeroth and first-order terms (higher-order terms will vanish when we take the limit as $\delta \mapsto 0$). To perform this expansion, we use the assumption that this distribution is expressed in terms of gaussians. As $\delta \mapsto 0$, these gaussians will tend toward delta functions, and the derivatives of gaussians will tend toward derivatives of delta functions (thereby simplifying the expansion). This derivation can be justified rigorously, playing detailed attention to convergence issues, by the use of distribution theory or the application of Itô calculus. If we expand $G(\vec{v}(\vec{x}, t+\delta) - \vec{v}'(\vec{x}', t) : \delta\Sigma_{\vec{v}})$ about $\delta = 0$, we find that the zeroth-order term is a Dirac delta function with argument $\vec{v}(\vec{x}, t+\delta) - \vec{v}'(\vec{x}', t)$. This term can therefore be integrated out. The derivative with respect to δ will effectively correspond to differentiating the gaussian with respect to the covariance. By standard properties of the gaussian, this will be equivalent to a second-order spatial derivative. We will get similar terms if we expand $G(\vec{x} - \vec{x}' - \delta\vec{v}'(\vec{x}', t) : \delta\Sigma_{\vec{x}})$, but we will also get an additional drift term from differentiating the $\delta\vec{v}'(\vec{x}', t)$ argument. In addition, we will get other terms from the denominator of equation A.1. (These are required to normalize the distributions and are nonstandard. They are needed because we have a differential equation for a set of interacting probability distributions while the standard Kolmogorov's equation is for a single probability distribution.) We collect all these zeroth- and first-order terms in the expansion and substitute into equation A.3. We can then evaluate the integrals using known properties of the delta functions. After some algebra, these integrals yield our variant of Kolmogorov's equation:

$$\begin{aligned}\frac{\partial}{\partial t} P(\vec{v}(\vec{x}, t)) = \frac{1}{2} \left\{ \frac{\partial^T}{\partial \vec{v}} \Sigma_{\vec{v}} \frac{\partial}{\partial \vec{v}} \right\} P(\vec{v}(\vec{x}, t)) + \frac{1}{2} \left\{ \frac{\partial^T}{\partial \vec{x}} \Sigma_{\vec{x}} \frac{\partial}{\partial \vec{x}} \right\} P(\vec{v}(\vec{x}, t)) \\ - \vec{v} \cdot \frac{\partial}{\partial \vec{x}} P(\vec{v}(\vec{x}, t)) + \frac{1}{|V|} \frac{\partial}{\partial \vec{x}} \int \vec{v} P(\vec{v}(\vec{x}, t)) d\vec{v}, \end{aligned} \tag{A.3}$$

where $\{\frac{\partial^T}{\partial \vec{x}} \Sigma_{\vec{x}} \frac{\partial}{\partial \vec{x}}\}$ and $\{\frac{\partial^T}{\partial \vec{v}} \Sigma_{\vec{v}} \frac{\partial}{\partial \vec{v}}\}$ are scalar differential operators, and $|V|$ is the volume of velocity space $\int d\vec{v}$.

Acknowledgments

This work was supported by AFOSR grant F49620-95-1-0265 to A. L. Y. and N. M. G. Further support to A. L. Y. came from ARPA and the Office of Naval Research, grant number N00014-95-1-1022. In addition, N.M.G. received support from grants by the National Eye Institute (EY08921 and EY11170) and the William A. Kettlewell chair. Finally, we thank the National Eye Institute for a core grant to Smith-Kettlewell (EY06883). We appreciate useful comments from anonymous referees.

References

Adelson, E. H., & Bergen, J. R. (1985). Spatio-temporal energy models for the perception of motion. *J. Opt. Soc. A., A 2*, 284–299.

Anstis, S. M., & Ramachandran, V. S. (1987). Visual inertia in apparent motion. *Vision Res., 27*, 755–764.

Ascher, D. (1998). *Human visual speed perception: Psychophysics and modeling*. Unpublished doctoral dissertation, Brown University.

Bar-Shalom, Y., & Fortmann, T. E. (1988). *Tracking and data association*. Orlando, FL: Academic Press.

Blake, A., & Yuille, A. L. (eds.). (1992). *Active vision*. Cambridge, MA: MIT Press.

Burr, D. C., Ross, J., & Morrone, M. C. (1986). Seeing objects in motion. *Proc. Royal Soc. Lond. B, 227*, 249–265.

Chin, T. M., Karl, W. C., & Willsky, A. S. (1994). Probabilistic and sequential computation of optical flow using temporal coherence. *IEEE Trans. Image Proc., 3*, 773–788.

Clark, J. J., & Yuille, A. L. (1990). *Data fusion for sensory information processing systems.* Norwell, MA: Kluwer.

Courant, R., & Hilbert, D. (1953). *Methods of mathematical physics* (Vol. 1). New York: Interscience.

Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. New York: Wiley.

Gottsdanker, R. M. (1956). The ability of human operators to detect acceleration of target motion. *Psych. Bull., 53*, 477–487.

Grzywacz, N. M., Amthor, F. R., & Mistler, L. A. (1990). Applicability of quadratic and threshold models to motion discrimination in the rabbit retina. *Biol. Cybern., 64*, 41–49.

Grzywacz, N. M. & Koch, C. (1987). Functional properties of models for direction selectivity in the retina. *Synapse, 1*, 417–434.

Grzywacz, N. M., Smith, J. A., & Yuille, A. L. (1989). A computational framework for visual motion's spatial and temporal coherence. In *Proc. IEEE Workshop on Visual Motion*. Irvine, CA.

Grzywacz, N. M., Watamaniuk, S. N. J., & McKee, S. P. (1995). Temporal coherence theory for the detection and measurement of visual motion. *Vision Res., 35*, 3183–3203.

Grzywacz, N. M., & Yuille, A. L. (1990). A model for the estimate of local image velocity by cells in the visual cortex. *Proceedings of the Royal Society of London B, 239*, 129–161.

Hassenstein, B., & Reichardt, W. E. (1956). Systemtheoretische analyse der zeit-, reihenfolgen- und vorzeichenauswertung bei der bewegungsperzeption des russelkafers. *Chlorophanus Z. Naturforsch., 11b*, 513–524.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neurosci., 9*, 181–197.

Ho, Y-C., & Lee, R. C. K. (1964). A Bayesian approach to problems in stochastic estimation and control. *IEEE Trans. on Automatic Control, 9*, 333–339.

Holub, R. A. & Morton-Gibson, M. (1981). Response of visual cortical neurons of the cat to moving sinusoidal gratings: Response-contrast functions and spatiotemporal integration *J. Neurophysiol., 46*, 1244–1259.

Huber, P. J. (1981). *Robust statistics*. New York: Wiley.

Isard, M., & Blake A. (1996). Contour tracking by stochastic propagation of conditional density. In *Proc. Europ. Conf. Comput. Vision* (pp. 343–356). Cambridge, UK.

Isard, M., & Blake, A. (1998). A mixed-state condensation tracker with automatic model-switching. *Proc. Int. Conf. Comp. Vis.* (pp. 107–112). Mumbai, India.

Jazwinski, A. H. (1970). *Stochastic processes and filtering theory*. Orlando, FL: Academic Press.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Trans. ASME, D: J. Basic Engineering, 82*, 35–45.

Kanizsa, G. (1979). *Organization in vision: Essays on gestalt perception*. New York: Praeger.

Kulikowski, J. J., & Tolhurst, D. J. (1973). Psychophysical evidence for sustained and transient detectors in human vision. *Journal of Physiology, 232*, 519–548.

Matthies, L., Kanade, T., & Szeliski, R. (1989). Kalman filter-based algorithms for estimating depth from image sequences. *Int. J. Comput. Vision, 3*, 209–236.

Maunsell, J. H. R., & Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. In W. M. Cowan, E. M. Shooter, C. F. Stevens, & R. F. Thompson, (Eds.), *Annual reviews of neuroscience* (Vol. 10, pp. 363–401). Palo Alto, CA: Annual Reviews.

McKee, S. P., Silverman, G. H., & Nakayama, K. (1986). Precise velocity discrimination despite random variations in temporal frequency and contrast. *Vision Res., 26*, 609–619.

Mel, B. W. (1993). Synaptic integration in an excitable dendritic tree. *J. Neurophysiol., 70*, 1086–1101.

Mel, B. W., Ruderman, D. L., & Archie, K. A. (1988). Translation-invariant orientation tuning in visual complex cells could derive from intradendritic computations. *J. Neurosci., 1*, 4325–34.

Merigan, W. H., & Maunsell, J.H. (1993). How parallel are the primate visual pathways? In W. M. Cowan, E. M. Shooter, C. F. Stevens, & R. F. Thompson

(Eds.), *Annual reviews of neuroscience* (Vol. 16, pp. 369–4021). Palo Alto, CA: Annual Reviews.

Nakayama, K. (1985). Biological image motion processing: A review. *Vision Res., 25*, 625–660.

Nishida, S., & Johnston, A. (1999). Influence of motion signals on the perceived position of spatial pattern. *Nature, 397*, 610–612.

Poggio, T. & Reichardt, W. E. (1973). Considerations on models of movement detection. *Kybernetik, 13*, 223–227.

Poggio, T., & Reichardt, W. E. (1976). Visual control of orientation behaviour in the fly: Part II: Towards the underlying neural interactions. *Q. Rev. Biophys., 9*, 377–438.

Ramachandran, V. S., & Anstis, S. M. (1983). Extrapolation of motion path in human visual perception. *Vision Res., 23*, 83–85.

Rao, R. P. N., & Ballard, D. H. (1997). Dynamic model of visual recognition predicts neural response properties of the visual cortex. *Neural Computation, 9*, 721–734.

Reichardt, W. (1961). Autocorrelation, a principle for the evaluation of sensory information by the nervous system. In Rosenblith (Ed.), *Sensory communication*. New York: Wiley.

Snowden, R. J., & Braddick, O. J. (1989). Extension of displacement limits in multiple-exposure sequences of apparent motion. *Vision Res., 29*, 1777–1787.

Torre, V., & Poggio, T. (1978). A synaptic mechanism possibly underlying directional selectivity to motion. *Proc. R. Soc. Lond. B, 202*, 409–416.

van Santen, J. P. H., & Sperling, G. (1984). A temporal covariance model of motion perception *J. Opt. Soc. A., A 1*, 451–473.

Verghese, P., Watamaniuk, S. N. J., McKee, S. P., & Grzywacz, N. M. (1999). Local motion detectors cannot account for the detectability of an extended motion trajectory in noise. *Vision Res., 39*, 19–30.

Wang, J. Y. A., & Adelson, E. H. (1993). Layered representation for motion analysis. In *Proc. IEEE Computer Sociey Conference on Computer Vision and Patern Recognition 1993* (pp. 361–366). New York.

Watamaniuk, S. N. J., & McKee, S. P. (1995). Seeing motion behind occluders. *Nature, 377*, 729–730.

Watamaniuk, S. N. J., McKee, S. P., & Grzywacz, N. M. (1994). Detecting a trajectory embedded in random-direction motion noise. *Vision Res., 35*, 65–77.

Watamaniuk, S. N. J., Sekuler, R., & Williams, D. W. (1989). Direction perception in complex dynamic displays: The integration of direction information. *Vision Research, 29*, 47–59.

Watson, A. B., & Ahumada, A. J. (1985). Model of human visual motion sensing. *J. Opt. Soc. A, A 2*, 322–342.

Welch, L., Macleod, D.I., & McKee, S. P. (1997). Motion interference: Perturbing perceived direction. *Vision Res, 37*, 2725–2736.

Werkhoven, P., Snippe, H. P., & Toet, A. (1992). Visual processing of optic acceleration. *Vision Res., 32*, 2313–2329.

Williams, L. R., & Jacobs, D. W. (1997a). Local parallel computation of stochastic completion fields. *Neural Comp., 9*, 859–881.

Williams, L. R., & Jacobs, D. W. (1997b). Stochastic completion fields: A neural

model of illusory contour shape and salience. *Neural Comp., 9*, 837–858.
Yuille, A. L., Burgi, P.-Y., & Grzywacz, N. M. (1998). Visual motion estimation and prediction: A probabilistic network model for temporal coherence. In *Proceedings of the Sixth International Conference on Computer Vision* (pp. 973–978). New York: IEEE Computer Society.
Yuille, A. L., & Grzywacz, N. M. (1988). A computational theory for the perception of coherent visual motion. *Nature, 333*, 71–74.
Yuille, A. L. & Grzywacz, N. M. (1989). A mathematical analysis of the motion coherence theory. *International Journal of Computer Vision, 3*, 155–175.
Yuille, A. L., & Grzywacz, N. M. (1998). A theoretical framework for visual motion. In T. Watanabe (Ed.), *High-level motion processing: Computational, neurobiological, and psychophysical perspectives*. Cambridge, MA: MIT Press.

Received November 2, 1998; accepted August 27, 1999.